Modelling Conditional Covariances with Orthogonal Factor Models

by

Tracy Jensen

Submitted to the Department of Statistical Sciences in partial fulfilment of the requirements for the degree of

Masters of Science in Mathematics of Finance
at the
UNIVERSITY OF CAPE TOWN

April 24, 2011

Supervisor: Prof. Linda Haines

Abstract

The recent sub prime crisis has resulted in an increased focus on risk management and monitoring in the financial industry. One of the essential components of risk management and monitoring is a reliable ex-ante covariance matrix of various financial time series. Therefore a reliable model which can handle a large number of time series is required to calculate an ex-ante or conditional covariance matrix. Some of the more common models used for this purpose are multivariate GARCH models. However, their use is restricted in practice as they are typically limited to modelling only a small number of time series.

A group of conditional covariance models which overcome this limitation are orthogonal factor models. Three specific orthogonal factor models are investigated in this thesis, namely the O-GARCH, O-EWMA and O-SV models. Moreover, a number of adjustments which can be made to the data and various model output are examined. Finally the suitability of the method used to calculate the factors is investigated.

The models are fitted to two different types of data. The first data set contains the returns of seven shares listed on the Johannesburg Stock Exchange and the second contains the exchange rate returns of five major currencies with the South African Rand. The three models of interest and the various adjustments are investigated in the context of these two data sets.

It is very difficult to make any definite conclusions from the results as it is not appropriate to assume that the results would generally be true for any financial data set. Although the results cannot be used to draw any definite conclusions, they do give some indications which can be further tested. For example, the results suggest that overall one of the three models is preferable, namely the O-SV model.

I hereby grant the University of Cape Town permission to copy and disseminate this work, or part thereof, for the purposes of study and research.

Plagiarism Declaration

1. This dissertation is my own work. It has not been submitted before for any degree or examination to any other University.

2. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

3. Each significant contribution to, and quotation in, this dissertation from the work of other people has been cited and reference.

Signature

Date

Acknowledgements

I would like to thank my supervisor, Professor Haines for all her assistance, guidance and patience.

I would also like to thank my parents for all their support during this time.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Risk management and monitoring have become increasingly important in the financial industry, especially after the recent sub prime crisis. This includes internal monitoring, regulatory requirements and client requests. One of the essential components of risk management is a reliable ex-ante covariance matrix of the relevant asset returns. An ex-ante measure is essential because a forward looking measure is preferable to a historic measure as historic risk is not necessarily a good predictor of future risk. A few of the purposes of the covariance matrix in risk management and other applications in finance are briefly discussed.

The ex-ante covariance matrix is required to calculate the ex-ante tracking error of a fund. This is often requested by clients and it may also be specified in a fund's mandate that the tracking error must remain within a certain range to ensure that the level of risk in the fund is appropriate. Another purpose is diversification, whereby an investor can use the covariance matrix to determine which assets may help to lower the overall risk of a portfolio. Furthermore the covariance matrix may be used to assist with asset selection, where investors or fund managers need to decide on an appropriate compromise between the returns and the risks. In this context the variance of the portfolio's returns is commonly used to measure risk. Another common measure of risk is Value-at-Risk (VaR) which is often calculated assuming normality of the returns, in which case the covariance matrix of asset returns is required to compute the variance of a fund. Besides these purposes mentioned there are many more, such as asset pricing, pairs trading, hedging and so forth.

Moreover the need to calculate a covariance matrix is not restricted to that of asset returns. The covariance matrix of some other quantity may also be necessary, for example exchange rates returns. Even though an exchange rate is not an asset in and of itself, it affects the value of a portfolio which holds assets in a variety of currencies. Therefore the correlations of the relevant exchange rates are important. For the the above reasons, and many others, it is vitally important for stake holders in the financial industry to model the variance and correlation of the returns of various financial quantities.

It is common practice to use a multivariate GARCH (generalised autoregressive conditional heteroscedasticity) or factor model to determine the ex-ante or conditional covariance matrix. In some cases a multivariate GARCH model may also be a factor model. A recent review of the multivariate GARCH and factor models commonly used in practice to model covariances are given in Bauwens et al. (2006). The main models discussed include, inter alia, the vectorised GARCH (VEC) of Bollerslev et al. (1988), the BEKK of Baba, Engle, Kraft and Kroner (Engle and Kroner, 1995), the constant conditional correlation (CCC) of Bollerslev (1988), the dynamic conditional correlation (DCC) models of Tse and Tsui (2002) and Engle (2002), the full-factor multivariate GARCH of Vrontos et al. (2002), the orthogonal GARCH (O-GARCH) model of Ding (1994) and Chibumba and Alexander (1996) and the generalised orthogonal GARCH (GO-GARCH) model of van der Weide (2002). Another important model that was introduced after 2006, models the multivariate volatilities via conditionally uncorrelated components (CUCs) (Fan et al., 2008).

1.2 Rationale for using an Orthogonal Factor Model

There are two major problems with many meaningful multivariate GARCH models. The first is the large number of parameters which need to be estimated and this number typically increases quadratically with the number of time series being modelled. Due to the large number of parameters, the likelihood function will be fairly flat which results in convergence problems in the optimisation routines that make it difficult to accurately estimate the parameters (Fan et al., 2008). Hence many of these models are only used in circumstances where there are only a few time series being modelled and this greatly restricts their use in practice. The second major problem is that many of the multivariate GARCH models require constraints imposed on them to ensure that the covariance matrix generated by the model is in fact positive semi-definite. This is a necessary but not a sufficient condition for a matrix to be considered a covariance matrix.

A multivariate GARCH and factor model which avoids both of the above problems is the orthogonal GARCH model (O-GARCH). This model was first suggested by Ding (1994) in his PhD thesis and then developed by Chibumba and Alexander (1996). A more general form of this model is considered in this thesis, that is a class of orthogonal factor models.

Orthogonal factor models overcome the problem that only a few time series can be modelled simultaneously and the positive semi-definite restrictions. Additionally the orthogonal factors are constructed out of principal component scores which has the benefit of ease of implementation.

1.3 Aims of the Thesis

It is apparent that many papers written on covariance factor models have focussed on testing various methods of calculating the factors. However, only one method is used to construct the factors in this thesis, that of principal component scores. This approach is chosen due to its simplicity and relative ease of implementation which ensures its suitability for wide usage by the relevant market participants. Therefore as opposed to focussing on different methods of calculating the factors, the focus of this thesis is on finding alternative methods to improve the model fit. These methods involve investigating different ways of modelling the conditional variance of the factors and testing various methods of adjusting the data and model output. The reason for this focus is to retain the simple idea of principal components while still attempting to find other methods to improve on the model.

Therefore the aims of this thesis are:

1. To provide a detailed understanding of both the theory and steps necessary to implement an orthogonal factor model to ensure understanding and ease of implementation. This is necessary as a review of the literature suggests that the model steps of a general factor model have not been described in detail. In general, papers focus on a specific model

and even then the steps of the specific model are typically not described in sufficient detail to ensure easy implementation in practice.

2. Both van der Weide (2002) and Fan et al. (2008) suggest that the principal component method has a number of problems associated with it. Therefore one of the aims of this thesis is to try to ascertain whether this choice of factors has a negative impact on all the orthogonal factor models when fitted to the data sets in this thesis.

3. To analyse and compare three different orthogonal factor models and the various model adjustments in the context of South African data. This includes examining the reasonability of the results.

4. Additionally the models are fitted to two data sets to determine if there is much of a difference in the appropriateness of the models when fitted to different data types.

1.4 Outline of the Thesis

A broad overview of general multivariate factor models, the model form and notation are given in Chapter 2. Furthermore, this chapter briefly discusses a few of the commonly used factor models. Chapter 3 gives an overview of the theory behind a general orthogonal factor model and describes the three specific orthogonal models tested in this thesis in more detail. The various adjustments made to the data and model output are given in Chapter 4. Chapter 5 investigates the properties and characteristics of the two data sets used in this thesis. In Chapter 6 the results of the conditional mean models fitted to the data sets are analysed. Chapter 7 is devoted to the resulting model parameter estimates and their interpretation. Chapter 8 reviews the model results in their entirety, as well as comparing the different models and the various adjustments. In conclusion, Chapter 9 discusses the aims achieved in this thesis, as well as future work which is beyond the scope of this study.

Chapter 2

Introduction to Multivariate Factor Models

In this chapter a general factor model is described along with some of the more commonly used factor models. The main differences between these models are the methods used to construct the factors which give rise to the estimated covariance matrix.

2.1 Overview of the Model and Notation

In this thesis the factor models will be fitted to two different types of data. The first data set contains stock market prices and the second exchange rates. These time series generally display a trend over time so they are converted to log returns to remove the trend. However, even after the trend has been removed these log returns are typically non-stationary. This is because the volatilities and correlations change over time. Therefore in the present study these quantities are allowed to change over time by modelling the conditional volatilities and correlations where the conditional information is a data set containing all the log returns up to the current time point.

Hence to calculate the conditional volatilities or correlations the conditional means first need to be removed. In this thesis the conditional mean is estimated by fitting an autoregressive moving average (ARMA) model or vector autoregressive (VAR) model to each series of log returns. The fitted residuals of the ARMA or VAR model should have a zero conditional mean. These residuals can then be used as they are or, as suggested by Alexander (2003), Tsay (2005) and Fan et al. (2008), they can be standardised or whitened. However any two series of residuals (whether adjusted or not) will typically have non-zero conditional correlations so the conditional volatilities and correlations of all the series would need to be modelled together, that is a multivariate model is required. To overcome the difficulties of multivariate volatility modelling the residuals can be converted to factors where any pair of factors have zero conditional correlations over time. These factors may be standardised by multiplying each factor by a constant which means that the standardised factors retain the property of zero conditional correlations. This allows the conditional volatility of each series of standardised factors to be modelled separately with a univariate model. The standardised factor volatilities can then be reverse engineered to determine the conditional volatilities and correlations of the log returns, which are used to construct a conditional covariance matrix. This process is summarised in Figure 2.1 along with an introduction to some of the notation used.

2.2 Mean and Covariance of Returns

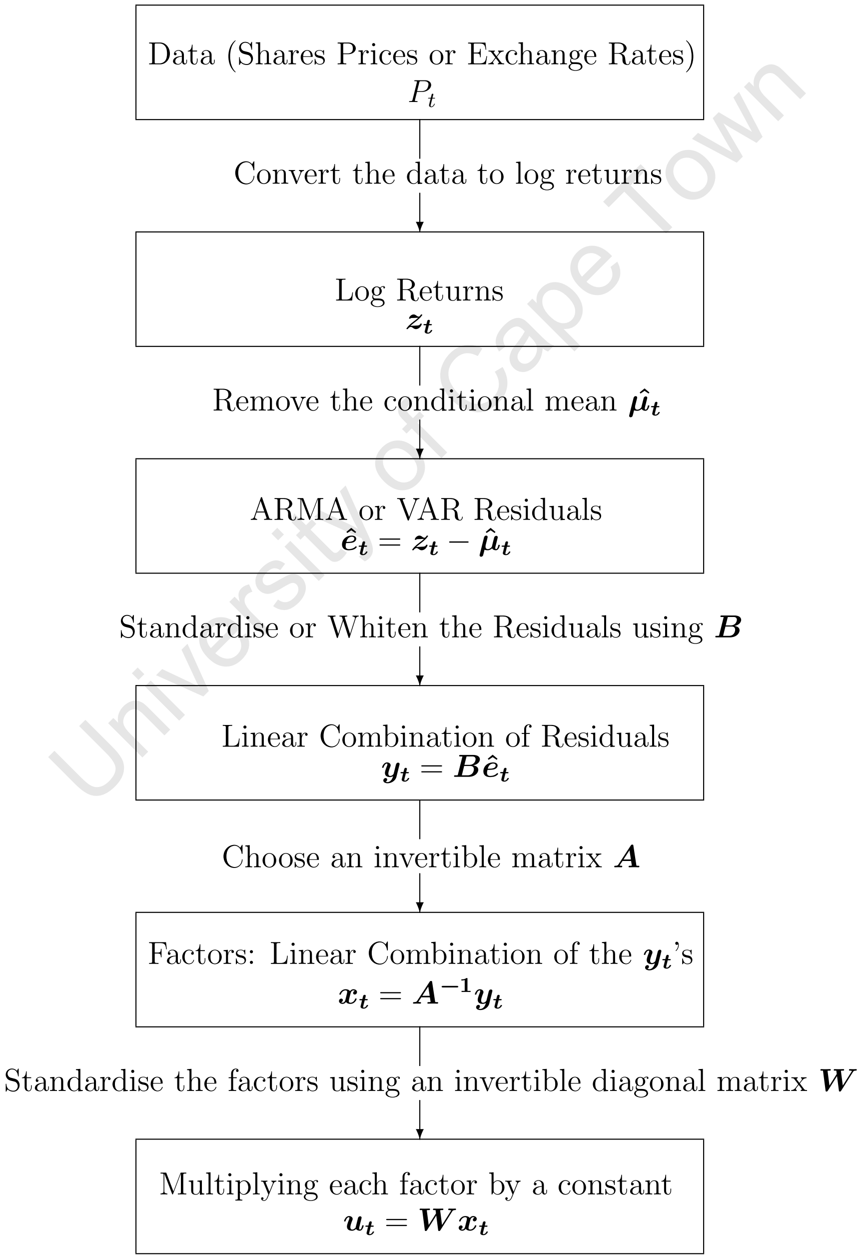
The modelling process and notation are now considered in greater detail. Let $\boldsymbol{z_t}$ be a column vector of the log returns of N assets (or some other quantity such as exchange rates) at time t with $t = 1, 2, \cdots, T$ such that $\boldsymbol{z_t} = (z_{1t}, z_{2t}, \cdots, z_{Nt})^T$ where z_{it} is the log return of the i^{th} series at time t. Although the log returns $\boldsymbol{z_t}$ are not stationary because the covariances are assumed to change over time, to simplify matters they are assumed to be stationary for the purposes of calculating the unconditional mean and covariance matrix. Therefore assuming that $\boldsymbol{z_t}$ is stationary let the unconditional population mean and covariance matrix be denoted as

$$\begin{aligned} E[\boldsymbol{z_t}] &= \boldsymbol{\mu} \qquad \text{an } N \times 1 \text{ vector} \\ Var[\boldsymbol{z_t}] &= \boldsymbol{S} \qquad \text{an } N \times N \text{ matrix} \end{aligned}$$

with the sample estimates of the mean and covariance matrix given by $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{S}}$ respectively.

However, the main focus of this thesis is on the conditional covariance matrix and therefore the information sets upon which the data are conditioned need to be defined. Let F_t be the information set available at time t (that is a σ-algebra generated by $\{\boldsymbol{z_t},\ \boldsymbol{z_{t-1}},\ \cdots,\ \boldsymbol{z_1}\}$). Let the conditional mean and

Figure 2.1: Flow Chart of the Modelling Process and Notation



covariance at time t be denoted as

$$\begin{aligned} E[\boldsymbol{z_t}|F_{t-1}] &= \boldsymbol{\mu_t} \\ Var[\boldsymbol{z_t}|F_{t-1}] &= \boldsymbol{S_t}. \end{aligned}$$

with the sample estimates of the conditional mean and covariance given by $\boldsymbol{\hat{\mu}_t}$ and $\boldsymbol{\hat{S}_t}$ respectively.

In theory the conditional mean of asset returns should be zero otherwise a trader could estimate the return at some time point in the future using the information today. If this were possible then arbitrage opportunities exist and markets would be considered to be inefficient. However in practice if a conditional mean is fitted to the data it may be statistically significant (i.e. not zero). This could be because some markets may not be as efficient as hypothesised (e.g. less liquid third world stock markets). Futhermore, regardless of the market, if the asset is thinly traded then it is likely to have a non zero sample conditional mean. However perhaps a more common reason for obtaining a non zero estimate of the conditional mean is that there may be a few extreme returns adjacent to each other which can induce a non zero conditional sample mean. Thus in practice there may be a statistically significant conditional mean and therefore for the purpose of calculating the conditional covariances this mean needs to be removed.

2.3 Mean and Covariance of the ARMA or VAR Residuals

There are a number of methods which can be used to estimate the conditional mean. For example an autoregressive moving average (ARMA) model could be fitted to each series of asset returns or a vector autoregressive (VAR) model fitted to all the series collectively. Therefore the conditional mean corrected series of log returns are given by

$$\boldsymbol{e_t} = \boldsymbol{z_t} - \boldsymbol{\mu_t} \qquad \text{an } N \times 1 \text{ vector}$$

which is essentially a vector containing the innovations of the ARMA or VAR model. However the actual residuals are never known and only the sample innovations can be observed as

$$\boldsymbol{\hat{e}_t} = \boldsymbol{z_t} - \boldsymbol{\hat{\mu}_t} \qquad \text{an } N \times 1 \text{ vector.}$$

These sample innovations $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ are a function of $\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$ which in turn depends on the estimated parameters in the ARMA or VAR model. Therefore the statistical properties of the sample innovations depend on the method used to fit the ARMA or VAR model. Hence the means and covariances of the theoretical residuals $\boldsymbol{e}_{\boldsymbol{t}}$ are considered instead of the sample residuals as these are not dependent on the method of fit. For practical purposes it is assumed that the sample residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ have the same conditional and unconditional means and covariances as the actual residuals $\boldsymbol{e}_{\boldsymbol{t}}$, which is not an unreasonable assumption to make.

The conditional means of the residuals $\boldsymbol{e}_{\boldsymbol{t}}$ are zero because the residuals are constructed by removing the conditional mean from $\boldsymbol{z}_{\boldsymbol{t}}$. Similarly the unconditional mean is also zero because $E[\boldsymbol{\mu}_{\boldsymbol{t}}] = \boldsymbol{\mu} = E[\boldsymbol{z}_{\boldsymbol{t}}]$.

The conditional covariance of the residuals $\boldsymbol{e}_{\boldsymbol{t}}$ is the same as the conditional covariance of the log returns $\boldsymbol{z}_{\boldsymbol{t}}$ because removing the conditional mean has no impact on the conditional covariance. However the unconditional covariance matrix will not be the same. So let $\boldsymbol{\Phi}$ be the unconditional covariance of $\boldsymbol{e}_{\boldsymbol{t}}$. Therefore assuming that the means and variances of $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ and $\boldsymbol{e}_{\boldsymbol{t}}$ are the same

$$\begin{aligned}
E[\hat{\boldsymbol{e}}_{\boldsymbol{t}}] &= \mathbf{0} && \text{an } N \times 1 \text{ vector} \\
Var[\hat{\boldsymbol{e}}_{\boldsymbol{t}}] &= \boldsymbol{\Phi} && \text{an } N \times N \text{ matrix} \\
E[\hat{\boldsymbol{e}}_{\boldsymbol{t}}|F_{t-1}] &= \mathbf{0} && \text{an } N \times 1 \text{ vector} \qquad (2.1) \\
Var[\hat{\boldsymbol{e}}_{\boldsymbol{t}}|F_{t-1}] &= \boldsymbol{S}_{\boldsymbol{t}} && \text{an } N \times N \text{ matrix} \qquad (2.2)
\end{aligned}$$

such that $\hat{\boldsymbol{\Phi}}$ is the sample estimate of $\boldsymbol{\Phi}$ and as previously mentioned $\hat{\boldsymbol{S}}_{\boldsymbol{t}}$ is the sample estimate of $\boldsymbol{S}_{\boldsymbol{t}}$. No symbols are required for the unconditional and conditional sample means as they will always be exactly zero by construction. It is obvious that the conditional sample mean will be zero because $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ is constructed by removing the conditional sample mean. However the reason that the unconditional sample mean is zero is because the ARMA or VAR model is fitted so that the sample residuals have an unconditional sample mean of zero.

2.4 Mean and Covariance of the Adjusted ARMA or VAR Residuals

Before the residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ are converted to factors they can be standardised or whitened as suggested by Alexander (2003) and Fan et al. (2008). This is done by adjusting the residuals by the unconditional covariance matrix $\hat{\boldsymbol{\Phi}}$ such that the adjusted residuals can be represented as a linear combination of the unadjusted residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$. These adjustments are discussed in more detail in Section 4.2. Therefore let $\boldsymbol{y_t} = (y_{1t}, y_{2t}, \cdots, y_{Nt})^T$ be used to represent the adjusted residuals such that $\boldsymbol{y_t}$ is a linear combination of the ARMA or VAR sample innovations $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ with

$$\begin{aligned} \boldsymbol{y_t} &= \boldsymbol{B}\hat{\boldsymbol{e}}_{\boldsymbol{t}} \\ &= \boldsymbol{B}(\boldsymbol{z_t} - \hat{\boldsymbol{\mu}}_{\boldsymbol{t}}) \qquad \text{an } N \times 1 \text{ vector} \end{aligned} \tag{2.3}$$

where $\boldsymbol{B}$ is an $N \times N$ invertible matrix. In the case where $\boldsymbol{B}$ is the identity matrix $\boldsymbol{y_t}$ will simply be the unadjusted residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$.

Some statistical properties of the adjusted residuals should be noted. Both the conditional and unconditional mean of $\boldsymbol{y_t}$ are zero because any linear combination of the sample innovations $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ will have a zero conditional and unconditional mean since the innovations $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ themselves have a zero conditional and unconditional mean. The same reasoning applies to the unconditional and conditional sample means.

On the other hand the unconditional and conditional covariance matrices of $\boldsymbol{y_t}$ and $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ will differ unless $\boldsymbol{B}$ is the identity matrix. Let $\boldsymbol{V}$ be the unconditional covariance of $\boldsymbol{y_t}$ and let the sample estimate of the covariance be given by $\hat{\boldsymbol{V}}$. Hence

$$\begin{aligned} E[\boldsymbol{y_t}] &= \boldsymbol{0} \qquad \text{an } N \times 1 \text{ vector} \\ Var[\boldsymbol{y_t}] &= \boldsymbol{V} \qquad \text{an } N \times N \text{ matrix} \\ &= \boldsymbol{B\Phi B^T}. \end{aligned}$$

Similarly let $\boldsymbol{V_t}$ be the conditional covariance of $\boldsymbol{y_t}$ so that

$$\begin{aligned} E[\boldsymbol{y_t}|F_{t-1}] &= \boldsymbol{0} \qquad \text{an } N \times 1 \text{ vector} \\ Var[\boldsymbol{y_t}|F_{t-1}] &= \boldsymbol{V_t} \qquad \text{an } N \times N \text{ matrix} \\ &= \boldsymbol{B S_t B^T}. \end{aligned}$$

where the sample estimate of the covariance $\boldsymbol{V_t}$ is given by $\hat{\boldsymbol{V}}_{\boldsymbol{t}}$.

Therefore the data are now in a suitable form to be converted to factors. This is because the general drift has been removed, the conditional mean is zero and the residuals have been standardised.

2.5 Factor Model

In the factor models adopted in this thesis it is assumed that a series of unobserved variables or factors exist which are related to the adjusted residuals $\boldsymbol{y_t}$ in some respect. The adjusted residuals $\boldsymbol{y_t}$ are assumed to be a linear combination of the latent variables or factors $\boldsymbol{x_t}$ with $\boldsymbol{x_t} = (x_{1t}, x_{2t}, \cdots, x_{Nt})^T$ where x_{it} is the i^{th} series of latent factors at time t. The adjusted residuals and factors are connected via an invertible matrix $\boldsymbol{A}$ such that

$$\boldsymbol{y_t} = \boldsymbol{A}\boldsymbol{x_t} \tag{2.4}$$

for times $t = 1, 2, \cdots, T$. Therefore $\boldsymbol{x_t} = \boldsymbol{A^{-1}}\boldsymbol{y_t}$ which implies that

$$\begin{aligned} E[\boldsymbol{x_t}|F_{t-1}] &= E[\boldsymbol{A^{-1}}\boldsymbol{y_t}|F_{t-1}] \\ &= \boldsymbol{0} \qquad\qquad \text{since } E[\boldsymbol{y_t}|F_{t-1}] = \boldsymbol{0}, \end{aligned}$$

that is the conditional mean of each of the factors is zero. Since the conditional mean of $\boldsymbol{y_t}$ is zero, the conditional mean of $\boldsymbol{x_t}$ will also be zero because $\boldsymbol{x_t}$ and $\boldsymbol{y_t}$ are linearly related. As a result the unconditional mean of $\boldsymbol{x_t}$ is zero because $E[E[\boldsymbol{x_t}|F_{t-1}]] = E[\boldsymbol{x_t}]$. Alternatively one could reason that the unconditional mean of $\boldsymbol{x_t}$ is zero because $E[\boldsymbol{x_t}] = E[\boldsymbol{A^{-1}}\boldsymbol{y_t}] = \boldsymbol{0}$.

Similarly the conditional covariance of the factors $\boldsymbol{x_t}$ can be represented in terms of the conditional covariance of the adjusted residuals $\boldsymbol{y_t}$. Let $\boldsymbol{D_t}$ be the conditional covariance matrix of $\boldsymbol{x_t}$ so that

$$\begin{aligned} Var[\boldsymbol{x_t}|F_{t-1}] &= \boldsymbol{D_t} \\ &= Var[\boldsymbol{A^{-1}}\boldsymbol{y_t}|F_{t-1}] \\ &= \boldsymbol{A}^{-1}\,\boldsymbol{V_t}\,(\boldsymbol{A}^T)^{-1} \qquad\qquad \text{an } N \times N \text{ matrix} \end{aligned}$$

and let $\boldsymbol{\hat{D}_t}$ be the sample estimate of $\boldsymbol{D_t}$. The unconditional covariance matrix of the factors $\boldsymbol{x_t}$ will be discussed in more detail at a later stage within the specific context of an orthogonal factor model.

Recall that a requirement of the factors $\boldsymbol{x_t}$ is that they are constructed in such a manner that any pair of factors are conditionally uncorrelated. This is to allow each factor to be modelled with a univariate model instead of modelling the data collectively with a multivariate model. By conditionally uncorrelated it is meant that $Cov(x_{it}, x_{jt}|F_{t-1}) = 0$ for $i \neq j$ and for all times $t = 1, 2, \cdots, T$. This assumption implies that the off diagonal elements of the conditional covariance matrix $\boldsymbol{D_t}$ are zero. Therefore let the conditional covariance matrix $\boldsymbol{D_t} = diag\{\sigma_{it}^2\}$ where σ_{it}^2 is the conditional variance at time t of the i^{th} latent factor (i.e. $Var[x_{it}|F_{t-1}] = \sigma^2_{it}$).

A probability distribution of the factors is also required in order to model the conditional variances σ^2_{it}. However the distribution of the factors will depend on the distribution of the ARMA or VAR residuals as the factors are a linear combination of these residuals. It is common practice to assume that the residuals of an ARMA or VAR model are normally distributed. However, in this thesis it is assumed that the conditional distribution of the ARMA or VAR residuals are normal as it is the conditional distribution which is required in the factor models and not the unconditional distribution. Since the factors are a linear combination of the residuals they will also be normally distributed as a linear combination of normally distributed variables is also a normally distributed variable. This result can be represented as

$$\boldsymbol{x_t}|F_{t-1} \sim N(\boldsymbol{0}, \boldsymbol{D_t}). \tag{2.5}$$

Finally the factors $\boldsymbol{x_t}$ may be standardised using an invertible, diagonal matrix $\boldsymbol{W} = diag\{w_{ii}\}$. Since the matrix is diagonal, standardising the factors is equivalent to multiplying each factor by the same constant at every time point. Therefore let $\boldsymbol{u_t} = (u_{1t}, u_{2t}, \cdots, u_{Nt})^T$ where u_{it} is the i^{th} series of standardised latent factors at time t with $u_{it} = w_{ii}x_{it}$. Therefore both the unconditional and conditional mean of the standardised factors are zero and the conditional covariance matrix is

$$\begin{aligned} Var[\boldsymbol{u_t}|F_{t-1}] &= Var[\boldsymbol{W x_t}|F_{t-1}] \\ &= Var[\boldsymbol{W A^{-1} y_t}|F_{t-1}] \\ &= \boldsymbol{W A}^{-1}\, \boldsymbol{V_t}\, (\boldsymbol{A}^{-1})^T \boldsymbol{W}^T \\ &= \boldsymbol{W D_t W}^T \quad \text{an } N \times N \text{ matrix.} \end{aligned} \tag{2.6}$$

Therefore the standardised factors are modelled as

$$\boldsymbol{u_t}|F_{t-1} \sim N(\boldsymbol{0}, \boldsymbol{W D_t W}^T). \tag{2.7}$$

On the other hand the unconditional covariance matrix of $\boldsymbol{u_t}$ is not relevant to the model so it is not discussed.

Hence if a model is chosen for the conditional variances of the standardised factors $\boldsymbol{u_t}$ then an estimate of $\boldsymbol{D_t}$ can be obtained using the relationship in equation (2.6) to give

$$\boldsymbol{D_t} = (\boldsymbol{W}^{-1}) Var[\boldsymbol{u_t}|F_{t-1}] (\boldsymbol{W}^T)^{-1}. \tag{2.8}$$

Furthermore the covariance matrix $\boldsymbol{D_t}$ can be reversed engineered to determine the conditional covariance matrix of the log returns $\boldsymbol{z_t}$. Therefore $\boldsymbol{V_t}$ can be estimated in terms of $\boldsymbol{D_t}$ using the relationship in factor model (2.4) to give

$$\begin{aligned} \boldsymbol{V_t} &= Var[\boldsymbol{y_t}|F_{t-1}] \\ &= Var[\boldsymbol{A}\boldsymbol{x_t}|F_{t-1}] \\ &= \boldsymbol{A}\,\boldsymbol{D_t}\,\boldsymbol{A}^T. \end{aligned} \tag{2.9}$$

Consequently an estimate of $\boldsymbol{D_t}$ will provide an estimate for $\boldsymbol{V_t}$. However what is required is the conditional covariance matrix of the log returns $\boldsymbol{z_t}$ which is $\boldsymbol{S_t}$. $\boldsymbol{S_t}$ can be computed as a function of $\boldsymbol{V_t}$ by invoking the result

$$\begin{aligned} \boldsymbol{S_t} &= Var[\boldsymbol{z_t}|F_{t-1}] \\ &= Var[\boldsymbol{\hat{e}_t}|F_{t-1}] \\ &= Var[\boldsymbol{B^{-1}}\boldsymbol{y_t}|F_{t-1}] \qquad \text{using equation (2.3)} \\ &= \boldsymbol{B^{-1}}\,\boldsymbol{V_t}\,(\boldsymbol{B^{-1}})^T. \end{aligned}$$

Therefore $\boldsymbol{S_t}$ is related to $\boldsymbol{D_t}$ as follows

$$\begin{aligned} \boldsymbol{S_t} &= \boldsymbol{B^{-1}}\,\boldsymbol{V_t}\,(\boldsymbol{B^{-1}})^T \\ &= \boldsymbol{B^{-1}}\,[\boldsymbol{A}\,\boldsymbol{D_t}\,\boldsymbol{A}^T]\,(\boldsymbol{B^{-1}})^T \\ &= [\boldsymbol{B^{-1}}\boldsymbol{A}]\,\boldsymbol{D_t}\,[\boldsymbol{B^{-1}}\boldsymbol{A}]^T. \end{aligned}$$

Therefore the aim of estimating and forecasting $\boldsymbol{S_t}$ is achieved. The form of $\boldsymbol{S_t}$ above ensures that it is always positive semi-definite without the need for an external constraint to be imposed. This is because $\boldsymbol{D_t}$ is positive semi-definite and is multiplied on one side by a matrix and on the other side by the transpose of the same matrix. This property is an advantage because a number of multivariate volatility models require additional constraints to be imposed to guarantee that the covariance matrices are positive semi-definite e.g. RiskMetrics™ (Alexander, 2003).

Finally for reasons that will become apparent at a later stage, it is desirable to consider the data in one large matrix in addition to the vector notation used above. Recall that $\boldsymbol{y_t} = \boldsymbol{A}\boldsymbol{x_t}$, which can also be represented as $\boldsymbol{y_t^T} = \boldsymbol{x_t^T}\boldsymbol{A^T}$.

Since this equation is true for all times t this can be formulated as a matrix over all times $t = 1, 2, \cdots, T$ to give

$$\begin{bmatrix} \boldsymbol{y_1^T} \\ \boldsymbol{y_2^T} \\ \vdots \\ \boldsymbol{y_T^T} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x_1^T} \\ \boldsymbol{x_2^T} \\ \vdots \\ \boldsymbol{x_T^T} \end{bmatrix} \boldsymbol{A^T} \qquad \text{or} \qquad \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A^T} \tag{2.10}$$

where $\boldsymbol{Y}$ and $\boldsymbol{X}$ are $T \times N$ matrices such that $\boldsymbol{Y}^T = (\boldsymbol{y_1}, \boldsymbol{y_2}, \cdots, \boldsymbol{y_T})$ is analogous to a data matrix and $\boldsymbol{X}^T = (\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_T})$ is a matrix of latent factors. Therefore $\boldsymbol{Y}$ is a matrix for which each column represents one series of adjusted residuals from time $t = 1$ to $t = T$ and each row represents the adjusted residuals at one time point for series 1 to N.

2.6 Variants of the Factor Model

The factor models discussed in the previous section have omitted two key details. These two details are what distinguishes the various models from one another. The first detail is the form $\boldsymbol{A}$ will assume and the method used to estimate $\boldsymbol{A}$ and the second is the choice of model or process for the conditional variance of u_{it} (the i^{th} standardised latent factor at time t). The same process or model is fitted for each series but the parameters are allowed to vary for each i. For example a GARCH(1,1) may be fitted to each series of standardised factors with different parameters for each series.

Consequently there are four main variants of the factor model which will be considered. These four variants are:

1. Orthogonal Generalised Autoregressive Conditional Heteroscedasticity (O-GARCH) model of Chibumba and Alexander (1996)

2. Generalised O-GARCH (GO-GARCH) model of van der Weide (2002)

3. Full-Factor Multivariate GARCH model of Vrontos et al. (2002)

4. Conditionally Uncorrelated Components (CUC) model of Fan et al. (2008)

A brief discussion of the choice of $\boldsymbol{A}$ and conditional variance model, as well as the main advantages and disadvantages of each variant are discussed below.

2.6.1 O-GARCH Model

The O-GARCH model was first suggested by Ding (1994) and later extended by Chibumba and Alexander (1996). The model simply uses an orthogonal matrix $\boldsymbol{A}$, with the columns consisting of the eigenvectors of $\hat{\boldsymbol{V}}$, the unconditional sample covariance matrix of $\boldsymbol{y_t}$. They can be decomposed by considering the definition of principal components which give $x_{it} = \boldsymbol{a_i^T y_t}$ where $\boldsymbol{a_i}$ is the eigenvector of $\hat{\boldsymbol{V}}$ corresponding the the i^{th} largest eigenvalue or alternatively $\boldsymbol{a_i}$ is the i^{th} column of matrix $\boldsymbol{A}$. This converts into $\boldsymbol{x_t} = \boldsymbol{A^T y_t}$ which can be converted to factor equation (2.4) using the fact that the inverse of $\boldsymbol{A}$ is its transpose. Therefore $\boldsymbol{A}$ is invertible as required and its inverse is given by its transpose. This choice of $\boldsymbol{A}$ results in each column of $\boldsymbol{X}$ consisting of the principal component scores of $\boldsymbol{Y}$, which contains the conditional mean corrected data.

The second key feature of the O-GARCH model is that each series of standardised principal component scores (factors) are modelled using a GARCH(1,1) process. Hence an advantage of the O-GARCH model is that it is quick and easy to implement. This is because the principal component scores are easy to calculate, a GARCH(1,1) model is easy to fit and both concepts are well known.

However van der Weide (2002) points out that although O-GARCH is easy to implement it may be problematic because $\boldsymbol{A}$ may be difficult to identify. Simulation studies performed by van der Weide (2002) suggest that the O-GARCH estimate of $\boldsymbol{A}$ is generally not very close to the actual value of the orthogonal matrix $\boldsymbol{A}$ used to simulate the data. These simulations involved choosing a value for the matrix $\boldsymbol{A}$ and a value for the matrix of latent factors $\boldsymbol{X}$. The data matrix $\boldsymbol{Y}$ is calculated using model equation (2.10) (i.e. $\boldsymbol{Y} = \boldsymbol{X A^T}$). Then an O-GARCH model is fitted to the data $\boldsymbol{Y}$ to obtain an estimate of the matrices $\boldsymbol{A}$ and $\boldsymbol{X}$. These estimates are then compared to the values of $\boldsymbol{A}$ and $\boldsymbol{X}$ which were used to simulate the data matrix $\boldsymbol{Y}$. This process is repeated a number of times. van der Weide (2002) found that on average the estimated matrix $\boldsymbol{A}$ was fairly different to the matrix $\boldsymbol{A}$ used to simulate the data. Hence he concluded that there is a problem identifying the true value of $\boldsymbol{A}$. This problem is believed to be due to the fact that estimation is based solely on the unconditional information $\hat{\boldsymbol{V}}$ (van der Weide, 2002). Typically identification difficulties occur when the eigenvalues of $\hat{\boldsymbol{V}}$ are quite similar as this causes problems identifying $\boldsymbol{A}$ using the eigenvectors. To overcome the identification problem in the O-GARCH model van der Weide (2002) developed a model called GO-GARCH (Generalised

O-GARCH).

This leads to the second disadvantage which is the fact that if real world data could be represented by a factor model why should $\boldsymbol{A}$ be orthogonal. Therefore a disadvantage of O-GARCH is that $\boldsymbol{A}$ is constrained to be orthogonal, which represents only a small subset of all invertible matrices (van der Weide, 2002).

On the other hand an advantage of the O-GARCH model is that if the columns of the data matrix $\boldsymbol{Y}$ are closely correlated then only the first few principal components need to be included in the model (Alexander, 2003). In other words, a GARCH(1,1) model is only fitted to the first few standardised principal component scores. The remaining scores are excluded from the model as they are considered to be background noise. This is because in a highly correlated system the first few principal components usually account for a large percentage of the variability in the system. This concept will be discussed in more detail at a later stage.

However, if the data are not closely correlated then the data will need to be placed into groups so that the data within each group are closely correlated (Alexander, 2003). Each of these groups are modelled separately using an O-GARCH model. Therefore each group will have an estimate of $\boldsymbol{A}$ and $\boldsymbol{S_t}$ where the dimensions may be different for each group. These matrices from each group $\boldsymbol{S_t}$ can be used to construct part of one large conditional covariance matrix which includes all the groups. However the covariances of series in different groups will be missing as these have not been modelled. Alexander (2003) gives a suggestion for modelling these missing conditional correlations using the estimates of $\boldsymbol{A}$ from each group and by estimating the conditional correlations of the principal components across groups. However this large conditional covariance matrix $\boldsymbol{S_t}$ for all the groups is not guaranteed to be positive semi-definite. Hence the conditional covariance matrix of all the log returns $\boldsymbol{S_t}$ will not necessarily be positive semi-definite and will require certain constraints (Alexander, 2003) to be adhered to, to ensure that it is positive semi-definite. Having to group the data is therefore a major disadvantage as it removes the benefit of quick and easy implementation and the conditional covariances may have to be adjusted away from their initial estimates to ensure that $\boldsymbol{S_t}$ is positive semi-definite.

The last disadvantage of O-GARCH is that the factors are not necessarily conditionally uncorrelated. By conditionally uncorrelated it is meant that $Cov(x_{it}, x_{jt}|F_{t-1}) = 0$ for $i \neq j$ and for all time $t = 1, 2, \cdots, T$. This is because the factors are the principal component scores which are only unconditionally uncorrelated. However the factor model assumes that the factors comprising the vector $\boldsymbol{x_t}$ are conditionally uncorrelated. This assumption is evident in factor model (2.5) where $\boldsymbol{D_t}$ (the conditional covariance matrix of $\boldsymbol{x_t}$) has zero off diagonal elements. Consequently fitting a GARCH(1,1) model to each of the standardised factors separately inherently assumes zero conditional correlations. Instead a multivariate model which allows for non-zero conditional correlations should be fitted. The CUC model of Fan et al. (2008) overcomes this problem by constructing the matrix $\boldsymbol{A}$ such that $\boldsymbol{X}$ gives a matrix of conditionally uncorrelated components (CUCs).

2.6.2 GO-GARCH

The GO-GARCH model of van der Weide (2002) also assumes that each series of standardised factors are modelled using a GARCH(1,1) model but $\boldsymbol{A}$ is no longer restricted to being orthogonal but has the advantage of being chosen from the wider range of invertible matrices. In addition van der Weide (2002) advocates that GO-GARCH solves the identification problem of the O-GARCH model by constructing an invertible matrix $\boldsymbol{A}$ using the Singular Value Decomposition. Two of the matrices in the decomposition are estimated using the unconditional information ($\boldsymbol{V}$) and the third is estimated using the conditional information. Hence

$$\boldsymbol{A} = \boldsymbol{P\Lambda U}$$

where $\boldsymbol{P}$ is a matrix such that each column is an eigenvector of $\boldsymbol{V}$, $\boldsymbol{\Lambda}$ is a diagonal matrix with the diagonal elements given by the eigenvalues of $\hat{\boldsymbol{V}}$ and $\boldsymbol{U}$ is an orthogonal matrix estimated by maximising the log-likelihood which contains conditional information.

However a disadvantage of GO-GARCH is that it is more complex to implement than O-GARCH as some of the parameters in $\boldsymbol{A}$ are estimated simultaneously with the GARCH(1,1) parameters. This may result in convergence problems. In response to this problem Boswijk and van der Weide (2006) subsequently developed a new method for estimating the GO-GARCH parameters which is easier to implement and less likely to experience problems with convergence. Although this development improves the problems of GO-GARCH it is still more complicated to implement than the O-GARCH model.

Another disadvantage is that the comparative study in Fan et al. (2008) suggests that the GO-GARCH model provides a poor fit for the data sets considered in their paper. In fact the model resulted in a negative conditional correlation between the return of Intel and the S&P500, which is highly unlikely in practice.

2.6.3 Full-Factor Multivariate GARCH Model

The full-factor multivariate GARCH model of Vrontos et al. (2002) models each series of standardised factors using a GARCH(1,1) model and $\boldsymbol{A}$ is a triangular matrix. Consequently a disadvantage is that $\boldsymbol{A}$ is restricted to being triangular and invertible which represents a small set of all possible invertible matrices.

The matrix $\boldsymbol{A}$ is estimated by one of two methods. The first is using the maximum likelihood estimates and the second involves invoking the Bayesian paradigm and using Markov chain Monte Carlo algorithms. However the problem with using the maximum likelihood approach is that estimation of the triangular matrix is dependent upon the ordering of the series of returns. This ordering problem is overcome by using the Monte Carlo method. On the other hand an advantage of maximum likelihood is that the expected Fisher information matrix and the partial derivatives of the log likelihood are available in closed forms which simplifies calculations. Obviously the Monte Carlo approach does not have the advantage of closed forms. Yet another advantage with the maximum likelihood approach is that the Fisher scoring algorithm is not sensitive to the starting values.

2.6.4 CUC Model

The final model considered in this section is the CUC model of Fan et al. (2008). The CUC model overcomes the conditional correlation problem in O-GARCH by constructing an orthogonal matrix $\boldsymbol{A}$ such that $\boldsymbol{X}$ gives a matrix of conditionally uncorrelated components (CUCs). Therefore this model has the disadvantage of constraining $\boldsymbol{A}$ to be orthogonal, as did O-GARCH. Another problem is that the conditionally uncorrelated components do not always exist and their existence first needs to be tested before implementing the model.

Once $\boldsymbol{A}$ has been selected and the CUCs (factors) determined each series of standardised factors are modelled using an extended GARCH(1,1) model. The extended GARCH(1,1) model differs from the usual GARCH(1,1) model in that the conditional volatility at time t depends on the square of all the series of the standardised factors at time $t-1$ and not just on the square of standardised factor at time $t-1$ of the series which is being modelled.

Moreover the model is mathematically complicated and very detailed making it difficult to grasp for the practitioner. In addition due to its complexity it also makes the model difficult to implement, although Fan et al. (2008) do provide some software. Despite all the difficulties and disadvantages, the model appears to be very sound, as it accounts for the conditional correlations in an accurate and sophisticated manner.

2.6.5 Focus of the Model Variants

After reviewing the four variants of the factor models it should be evident that there is a common thread. All the variants model the standardised factors using a GARCH(1,1) model, or an extended GARCH(1,1) in the case of the CUC model. Hence the focus of these models has been on the choice of $\boldsymbol{A}$ and not on the choice of the conditional variance model. In contrast the focus of this thesis is on the conditional variance model and to some extent on other small adjustments such as the choice of $\boldsymbol{B}$, $\boldsymbol{W}$ and conditional mean model.

Chapter 3

Orthogonal Factor Models

In this thesis an orthogonal factor model is considered to be a factor model if it satisfies three requirements. The first is that it must satisfy equation (2.4), that is $\boldsymbol{y_t} = \boldsymbol{A}\boldsymbol{x_t}$. Secondly the model must satisfy equation (2.5), that is $\boldsymbol{x_t}|F_{t-1} \sim N(\boldsymbol{0}, \boldsymbol{D_t})$ with $\boldsymbol{D_t} = \boldsymbol{A^T V_t A}$. Finally the columns of $\boldsymbol{A}$ must consist of the eigenvectors of $\hat{\boldsymbol{V}}$, the sample covariance matrix of $\boldsymbol{y_t}$. Therefore $\boldsymbol{A}$ is orthogonal such that the specification of $\boldsymbol{A}$ is the same as in the O-GARCH model. To recap, this choice of $\boldsymbol{A}$ results in each column of $\boldsymbol{X}$ consisting of a series of the principal component scores of $\boldsymbol{Y}$. Consequently x_{it}, for $t = 1, 2, \cdots T$, are the principal component scores for the i^{th} asset or time series, with factor weights given by the columns of $\boldsymbol{A}$. Therefore let $\boldsymbol{a_i}$ be the i^{th} column of $\boldsymbol{A}$ so the equation $\boldsymbol{x_t} = \boldsymbol{A^T y_t}$ can be formulated as

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{NT}^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{a_1^T} \\ \boldsymbol{a_2^T} \\ \vdots \\ \boldsymbol{a_T^T} \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{NT}^T \end{bmatrix}.$$

Consequently it is evident that $x_{it} = \boldsymbol{a_i^T y_t}$ for $t = 1, 2, \cdots T$ which implies that $\boldsymbol{a_i}$ contains the factor weights of the i^{th} asset or time series.

As a result of the three basic specifications above orthogonal factor models have a number of advantages when used in practice. This is the reason that they are the focus of this thesis. The first practical advantage is that $\boldsymbol{A}$ is quick and easy to calculate. Secondly when there are missing observations or when some (but not all) of the assets are thinly traded, the covariance estimates of the returns are fairly reliable. This is important as many of the multivariate GARCH models cannot readily handle data of this nature, whereas orthogonal factor models only require reasonable estimates of the factor weights to obtain a model with an acceptable fit (Alexander, 2003).

This chapter therefore begins by discussing the general properties of orthogonal factor models and then proceeds with a discussion in more detail of three specific orthogonal factor models.

3.1 Important Properties of the Factors

A number of important properties of the factors arise from the fact that they are constructed as the principal component scores of $\boldsymbol{Y}$. Thus to fully understand the factors a brief definition of the eigenvalues and eigenvectors which are used to construct these principal component scores is necessary. Therefore let $\boldsymbol{\Lambda}$ be the diagonal matrix containing the eigenvalues of $\hat{\boldsymbol{V}}$ such that $\boldsymbol{\Lambda} = diag\{\lambda_i\}$ where λ_i is the i^{th} eigenvalue of $\hat{\boldsymbol{V}}$ and $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N$. Consequently $\boldsymbol{A}$ is a matrix where each column is an eigenvector of $\hat{\boldsymbol{V}}$ with the eigenvector in the first column corresponding to the eigenvalue λ_1 and the second to λ_2 and so forth. Thus the definition of the eigenvectors and eigenvalues of $\hat{\boldsymbol{V}}$ is

$$\hat{\boldsymbol{V}}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{\Lambda} \qquad \text{or} \qquad \boldsymbol{A^T}\hat{\boldsymbol{V}}\boldsymbol{A} = \boldsymbol{\Lambda}. \tag{3.1}$$

As a result one of the properties of the factors is that they are unconditionally uncorrelated because this is a property of the principal component scores. For the purposes of proving that the factors are unconditionally uncorrelated it is necessary to assume that $\boldsymbol{x_t}$ and $\boldsymbol{y_t}$ are stationary for $t = 1, 2, \cdots T$. Therefore $Var(\boldsymbol{x_t})$ is assumed to be the constant for $t = 1, 2, \cdots T$ and $Var(\boldsymbol{y_t})$ is assumed to be constant for $t = 1, 2, \cdots T$. Since $\boldsymbol{x_t}$ and $\boldsymbol{y_t}$ are related as $\boldsymbol{x_t} = \boldsymbol{A^T}\boldsymbol{y_t}$, the variances are related as $Var(\boldsymbol{x_t}) = \boldsymbol{A^T}Var(\boldsymbol{y_t})\boldsymbol{A}$. This relationship is also applicable if the unconditional covariance matrices are replaced with the estimates of these matrices. Thus let $\widehat{Var(\boldsymbol{x_t})}$ be the estimate of the unconditional covariance matrix of the factors and recall that $\hat{\boldsymbol{V}}$ is the unconditional covariance matrix of $\boldsymbol{y_t}$. Therefore

$$\begin{aligned} \widehat{Var(\boldsymbol{x_t})} &= \boldsymbol{A}^T\hat{\boldsymbol{V}}\boldsymbol{A} && \text{since } \boldsymbol{x_t} = \boldsymbol{A^T}\boldsymbol{y_t} \\ &= \boldsymbol{\Lambda} && \text{using definition (3.1).} \end{aligned}$$

Since $\mathbf{\Lambda}$ is a diagonal matrix the off diagonal elements of the sample estimate of the unconditional covariance matrix of $\boldsymbol{x_t}$ are zero for $t = 1, 2, \cdots T$. Therefore the factors are unconditionally uncorrelated.

Yet another property, which can be observed in the proof, is that the i^{th} eigenvalue is the sample variance of the i^{th} series of factors for $i = 1, 2, \cdots N$. Hence the proportion of volatility that the i^{th} factor contributes to the system as a whole is

$$\frac{\lambda_i}{\sum_{i=1}^{i=N} \lambda_i}.$$

With regard to this, Alexander (2003) suggests that when the log returns $\boldsymbol{z_t}$ are highly correlated then only a few series of factors will account for most of the variation in the system. For example if the first three series of factors represent 80% of the variation then

$$0.8 = \frac{\sum_{i=1}^{i=3} \lambda_i}{\sum_{i=1}^{i=N} \lambda_i}.$$

On the other hand in systems where the returns $\boldsymbol{z_t}$ have low correlations, the series should be grouped so that the series in any one group are highly correlated. This was discussed in detail in Section 2.6.1.

Due to the fact that a few factors often account for a large proportion of the volatility of the system Alexander (2003) advocates using only the first few series of principal component scores. There are two advantages to using only the first few series. Firstly fewer conditional variance models need to be fitted and therefore fewer parameters need to be estimated. Secondly using only the first few principal components will result in some of the background noise in the system being excluded, that is the volatility of the remaining series of principal component scores are excluded. Due to excessive noise being excluded from the system Alexander (2003) suggests that the estimates of the covariance of the returns $\boldsymbol{z_t}$ are more stable. However, whether this is more desirable depends on the purpose for which the correlation estimates are to be used. For example, a supposedly more accurate, non-smoothed estimate may be preferable for some purposes and a more stable estimate for others.

3.2 Model Details Using Only h Factors

Owing to the benefits of using only the first few series of principal component scores, as described in the previous section, the new model details are briefly given assuming that only the first h series are used. (If all series are included then $h = N$.) Let the new factor vector at time t be $\boldsymbol{x_t^{(h)}} = (x_{1t}, x_{2t}, \cdots, x_{ht})^T$ and let the new factor loadings be summarised in the matrix $\boldsymbol{A^{(h)}}$. Thus $\boldsymbol{A^{(h)}}$ is an $N \times h$ matrix with the columns of $\boldsymbol{A^{(h)}}$ consisting of the first h eigenvectors corresponding to the largest h eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_h$ of the covariance matrix $\hat{\boldsymbol{V}}$. However $\boldsymbol{A^{(h)}}$ is not orthogonal as it is not square. Although $(\boldsymbol{A^{(h)}})^T\boldsymbol{A^{(h)}}$ is a $h \times h$ identity matrix, $\boldsymbol{A^{(h)}}(\boldsymbol{A^{(h)}})^T$ is not an $N \times N$ identity matrix unless $h = N$.

Therefore the new factor vector is calculated as

$$\boldsymbol{x_t^{(h)}} = (\boldsymbol{A^{(h)}})^T\boldsymbol{y_t}$$

where the reworked model assumptions are

$$\boldsymbol{x_t^{(h)}}|F_{t-1} \sim N(\boldsymbol{0}, \boldsymbol{D_t^{(h)}}) \qquad \text{with}$$

$$\boldsymbol{D_t^{(h)}} = diag\{\sigma_{it}^2\} \qquad \text{for } i = 1, 2, \cdots, h.$$

As a result of the model having a new vector of factors there will also be a new vector of adjusted factors. Let this vector be $\boldsymbol{u_t^{(h)}} = (u_{1t}, u_{2t}, \cdots, u_{ht})^T$ which is related to the new factor vector $\boldsymbol{x_t^{(h)}}$ via an $h \times h$ diagonal, invertible matrix $\boldsymbol{W^{(h)}}$. Thus the relationship is $\boldsymbol{u_t^{(h)}} = \boldsymbol{W^{(h)}}\boldsymbol{x_t^{(h)}}$.

Previously, in the system where all N series of principal component scores are included, the conditional covariances of $\boldsymbol{y_t}$ and the factors $\boldsymbol{x_t}$ were related via equation (2.9). However when using only h series, the conditional covariances of $\boldsymbol{y_t}$ and $\boldsymbol{x_t^{(h)}}$ are related as

$$\begin{aligned} Var(\boldsymbol{y_t}|F_{t-1}) &= \boldsymbol{V_t} \\ &= \boldsymbol{A^{(h)}}\boldsymbol{D_t^{(h)}}(\boldsymbol{A^{(h)}})^T. \end{aligned}$$

Thus an estimate of $\boldsymbol{D_t^{(h)}}$ can be obtained by modelling the conditional volatility of each of the h series of standardised principal component scores. Once this estimate has been obtained the conditional covariance of the adjusted residuals $\boldsymbol{y_t}$ (that is $\hat{\boldsymbol{V_t}}$) can be obtained which is ultimately used to calculate $\hat{\boldsymbol{S}}_t$. The relationship between $\boldsymbol{V_t}$ and $\boldsymbol{S_t}$ is the same regardless of whether N or fewer series are included in the model. The same is true of the sample estimates $\hat{\boldsymbol{V}}_t$ and $\hat{\boldsymbol{S}}_t$.

From this point onwards all orthogonal factor models are assumed to use h of the possible N series of principal component scores and therefore the details above apply. However if $h = N$ then the details of this adjusted factor model will be identical to the initial factor model.

3.3 Overview of the Three Main Orthogonal Factor Models

An important detail of orthogonal factor models which need to be considered in detail is the choice of model for the conditional volatility of the standardised factors which is then used to estimate the covariance matrix $\boldsymbol{D_t^{(h)}}$. Three different conditional volatility models are evaluated in this thesis and are summarised as follows:

1. O-GARCH: Models the conditional volatility of the standardised factors using a GARCH(1,1) model.

2. Orthogonal Exponentially Weighted Moving Average (O-EWMA): Models the conditional volatility of the standardised factors using an IGARCH(1,1) model.

3. Orthogonal Stochastic Volatility (O-SV): Models the conditional volatility of the standardised factors using a stochastic volatility model introduced by Shephard (1994).

Each of these three models are now discussed in more detail.

3.4 Orthogonal GARCH

The O-GARCH model has already been discussed in some detail and is briefly revised here. The model assumes that each of the first h standardised principal component scores follow a GARCH(1,1) model. The representation of the GARCH(1,1) model is given by

$$\begin{aligned} u_{it} &= \sigma_{it}\,\epsilon_{it} && \text{with } \sigma_{it} \text{ and } \epsilon_{it} \text{ are independent} \\ \epsilon_{it} &\sim N(0,1) && \text{independent over } t \text{ and } i \\ \sigma_{it}^2 = Var(u_{it}|F_{t-1}) &= \alpha_{i,0} + \alpha_{i,1}\,u_{i,t-1}^2 + \beta_i\,\sigma_{i,t-1}^2 \end{aligned}$$

where $\alpha_{i,0}, \alpha_{i,1}$ and β_i are unknown parameters. The parameter constraints are $\alpha_{i,0} > 0$, $\alpha_{i,1}, \beta_i \geq 0$ and to ensure stationarity $0 < \alpha_{i,1} + \beta_i < 1$ for $i = 1, 2, \cdots, h$. The parameter $\alpha_{i,1}$ measures the extent of the markets reaction to the returns observed at the previous time step. Although this representation does not directly incorporate the returns, they are indirectly incorporated via the standardised principal component scores $\boldsymbol{u_t^{(h)}}$. The parameter β_i represents how persistent the volatility is (Alexander, 2000). Therefore the larger β_i is the greater the persistence.

In addition to interpreting the parameters, an interpretation of the term ϵ_{it} is required. Thus ϵ_{it} is an error term which is taken to be a white noise process over time $t = 1, 2, \cdots T$ for each $i = 1, 2, \cdots N$. The errors are assumed to follow a normal distribution because this is the most common assumption made when fitting a GARCH(1,1) and more importantly it is an assumption of the factor model in this thesis. The reason for choosing the standardised factors to be normally distributed was discussed in more detail in Section 2.5. Despite this choice being made it is possible to use other distributions for the errors ϵ_{it}, such as a Student's t distribution, but these will not be used in this thesis.

Beside interpreting the model, there are also some requirements to consider. The representation of the GARCH(1,1) model given above makes it evident that for a series to follow a GARCH(1,1) process the series must meet two requirements. These are that the unconditional mean is zero and that the conditional mean is zero, in other words the series is not autocorrelated. The first assumption is evident as $E[u_{it}] = E[\sigma_{it}\,\epsilon_{it}] = E[\sigma_{it}]\,E[\epsilon_{it}] = 0$ because σ_{it} and ϵ_{it} are independent and $E[\epsilon_{it}] = 0$. The second is evident in that $E[u_{it}|F_{t-1}] = E[\sigma_{it}\,\epsilon_{it}|F_{t-1}] = \sigma_{it}\,E[\epsilon_{it}|F_{t-1}] = \sigma_{it}\,E[\epsilon_{it}] = 0$ since σ_{it} is known given information F_{t-1}, the error term ϵ_{it} is independent of information F_{t-1} and $E[\epsilon_{it}] = 0$. Each series of standardised factors meet these two requirements (see Section 2.5).

3.4.1 Parameter Estimation

The parameters $\alpha_{i,0}, \alpha_{i,1}$ and β_i are estimated for each factor series individually, in other words without any information from any of the other series. The parameters for a given series of standardised factors are estimated by choosing the parameters which maximise the conditional log likelihood function. Hence the estimated parameters are maximum likelihood estimates. The log likelihood of the i^{th} series of standardised principal component scores is

$$LL(u_{i1}, \cdots, u_{iT}, \alpha_{i,0}, \alpha_{i,1}, \beta_i) = -Tlog(\sqrt{2\pi}) - \sum_{t=1}^{t=T} \left[log(\sigma_{it}) + \frac{1}{2} \frac{u_{it}^2}{\sigma_{it}^2} \right]$$

where σ_{it} is a function of the parameters $\alpha_{i,0}, \alpha_{i,1}$ and β_i. The maximum of the likelihood or log likelihood cannot be found analytically so a numerical technique must be invoked.

3.4.2 Forecasting the Conditional Volatility

Once the parameters have been estimated they can be used to forecast the conditional volatility given the information up to the current time. Let $\sigma_{i,t+k|t}^2$ be the forecast of the conditional variance at time $t+k$ given the information available at time t. For the case $k = 1$

$$\sigma_{i,t+1|t}^2 = \alpha_{i,0} + \alpha_{i,1}\, u_{it}^2 + \beta_i\, \sigma_{it}^2$$

and for $k > 1$

$$\sigma_{i,t+k|t}^2 = \alpha_{i,0} + (\alpha_{i,1} + \beta_i)\, \sigma_{i,t+k-1|t}^2$$

using the approximation that $u_{i,t+l}^2 = \sigma_{i,t+l|t}^2$ for $l \geq 1$ since $u_{i,t+l}^2$ is unknown given the information at time t.

Two advantages of O-GARCH are that the conditional variance forecasts converge to their long run mean and secondly that this forecast varies according to k, that is how far into the future the forecast is made (Alexander, 2000). This is important as the forecasts of some models remain constant after k is greater than a certain number.

3.5 Orthogonal Exponentially Weighted Moving Average Model

The orthogonal exponentially weighted moving average (O-EWMA) model is very similar to the O-GARCH model and was likewise developed by Carol

Alexander (2000). In fact, an O-EWMA model is an O-GARCH model with the restrictions that $\alpha_{i,0} = 0$ and $\alpha_{i,1} = 1 - \beta_i$. Therefore each of the series of standardised principal component scores are essentially modelled using an IGARCH(1,1) model without drift which is a special case of a GARCH(1,1) model. Hence IGARCH(1,1) has the disadvantage of being more constrained than a GARCH(1,1) model. Thus the IGARCH(1,1) model is represented as

$$\begin{aligned} u_{it} &= \sigma_{it}\,\epsilon_{it} \\ \epsilon_{it} &\sim N(0,1) \qquad \text{independent over time and } i \\ \sigma_{it}^2 = Var(u_{it}|F_{t-1}) &= \pi_i \sigma_{i,t-1}^2 + (1-\pi_i) u_{i,t-1}^2 \end{aligned} \tag{3.2}$$

where π_i is an unknown parameter, $0 < \pi_i < 1$ and $i = 1, 2, \cdots, h$.

As in the GARCH(1,1) model, the errors ϵ_{it} are assumed to follow a normal distribution. This is because it is a common assumption to make for an IGARCH(1,1) model and more importantly the standardised factors in this thesis are assumed to be normally distributed. Once again due to the fact that an IGARCH(1,1) model is simply a special case of a GARCH(1,1) model the requirement that the conditional and unconditional mean of each series of standardised factors are zero is applicable.

Equation (3.2) can be rewritten to further facilitate interpretation of the model. The conditional variance equation above can be represented as a sum by iteratively substituting an equation for $\sigma_{i,t-1}^2$ to give

$$\sigma_{it}^2 = \pi_i^{t-1}\sigma_{i1}^2 + (1-\pi_i)\sum_{k=1}^{t-1}[\pi_i^{k-1}u_{i,t-k}^2].$$

In this representation it is clear that an assumption about σ_{i1}^2 is required. One such possibility is to assume that σ_{i1}^2 is the unconditional variance of the i^{th} series of standardised principal component scores or to assume that it is u_{i1}^2.

Thus the representation above can be compared to the exponentially weighted moving average (EWMA) of the squares of the standardised scores which is

$$\sigma_{it}^2 = (1-\pi_i)\sum_{k=1}^{\infty}[\pi_i^{k-1}u_{i,t-k}^2]\,.$$

The two expressions for σ_{it}^2 above are very similar and in fact tend to the same value as $t \to \infty$. This is the reason for the model name O-EWMA. However if a drift term is included in the IGARCH(1,1) model then the two equations would not be as similar. This is one of the reasons for using an IGARCH(1,1) with no drift. A further reason is market convention.

With regards to market convention, one of the models frequently used to calculate the conditional covariances of a group of asset returns is RiskMetrics$^{\text{TM}}$ developed by J.P. Morgan and Reuters (1996). RiskMetrics$^{\text{TM}}$ models the conditional variance of each series of asset returns using a driftless IGARCH(1,1). The conditional covariance of each pair of asset returns is also modelled using a driftless IGARCH(1,1) except that instead of the square of the asset returns the cross product is used.

However the covariance matrix constructed from these estimates is not guaranteed to be positive semi-definite. Therefore to ensure that the covariance matrix is positive semi-definite the same smoothing constant (π_i) must be used for all the IGARCH(1,1) models (i.e. for each of the squares and cross products). As a result the same market reaction $(1-\pi_i)$ and persistence π_i are assumed for all the assets and cross products (Alexander, 2003). Therefore an advantage of O-EWMA over RiskMetrics$^{\text{TM}}$ is that the covariance matrix will always be positive semi-definite while still allowing a different smoothing constant to be used for each series of standardised principal component scores. Even if the same smoothing constant is used for all the standardised principal components, the smoothing constant will be different for each of the assets because the matrix $\boldsymbol{A}$ is used to convert the covariance of the standardised principal component scores to the covariance of the assets.

J.P. Morgan and Reuters (1996) suggest using a smoothing constant of 0.94 for daily returns and 0.97 for monthly returns based on the weighted average of the optimal smoothing constants over different asset classes. The optimal smoothing constant for each series is chosen by minimising the root mean of the predicted errors. RiskMetrics$^{\text{TM}}$ then take the weighted average of the optimal smoothing constant over 480 different time series to arrive at a smoothing constant of 0.94 for daily returns and 0.97 for monthly returns. These weights are calculated using the root mean squared of the prediction errors.

3.5.1 Parameter Estimation

The parameters are estimated separately for each series of standardised principal component scores. Considering a single series, the parameters are estimated by choosing the parameters which maximise the conditional log likelihood function. The log likelihood of the i^{th} series of standardised principal component scores is

$$LL(u_{i1}, \cdots, u_{iT}, \pi_i) = -Tlog(\sqrt{2\pi}) - \sum_{t=1}^{t=T} \left[log(\sigma_{it}) + \frac{1}{2}\frac{u_{it}^2}{\sigma_{it}^2} \right]$$

where σ_{it} is a function of the parameter π_i. The maximum of the likelihood or log likelihood cannot be found analytically so numerical techniques are required to estimate π_i.

Although the estimate of π_i can range between 0 and 1, Tsay (2005) suggests that usually a value between 0.9 and 1 is chosen when fitting an IGARCH(1,1) model directly to the asset returns, with 0.94 being the most common. However, for the data sets used in this thesis the maximum likelihood estimates of π_i are often considerably less than 0.9. This results in autocorrelation being induced into the errors ϵ_{it} which violates the important model assumption that the errors are a white noise process. These results are discussed in more detail in Section 7.2.2. Therefore the log likelihood was maximised with the constraint that the estimate for π_i must be such that the errors ϵ_{it} have no significant autocorrelation up to lag 5. In essence this means that the likelihood is maximised over a range which is smaller than $0 < \pi_i < 1$ and that this range depends on the autocorrelation of the errors ϵ_{it}. An error series was considered to have no significant autocorrelation up to lag 5 if the Ljung-Box Q-statistic was not significant at the 5% level. However if a series has a significant Q-statistic for any choice of π_i then the original maximum likelihood estimate for π_i is used.

3.5.2 Forecasting the Conditional Volatility

Once parameter estimates have been calculated these can be used to forecast the conditional volatility using the information up to that point in time. Let $\sigma^2_{i,t+k|t}$ be the forecast of the conditional variance at time $t+k$ given the information available at time t. For the case $k = 1$

$$\sigma^2_{i,t+1|t} = \pi_i \sigma^2_{it} + (1 - \pi_i) u^2_{it}$$

and for $k > 1$

$$\sigma^2_{i,t+k|t} = \sigma^2_{i,t+1|t}$$

using the approximation that $u^2_{i,t+l} = \sigma^2_{i,t+l|t}$ for $l \geq 1$ since $u^2_{i,t+l}$ is unknown given the information at time t.

One disadvantage of O-EWMA is that the covariance forecasts do not converge to their long run mean but are constant regardless of how far ahead the forecast is, that is they have a constant term structure (Alexander, 2000). This is evident in the fact that no matter how far into the future, that is how big k is, the conditional variance forecast is equal to the one step ahead forecast.

3.6 Orthogonal Stochastic Volatility Model

As noted earlier, one focus of this thesis is on modelling the conditional volatility of the standardised factors in a variety of manners. Hence a possible alternative to fitting a GARCH(1,1) model to each series of standardised principal component scores is to fit a stochastic volatility model to each series. Thus let the orthogonal factor model where each series of standardised principal component scores are assumed to follow a stochastic volatility model be termed Orthogonal Stochastic Volatility Model (O-SV). A name is given to the model because as far as the literature reviewed suggests, no paper has yet reported the implemented of the model. The stochastic volatility model used in this thesis is the Gaussian local scale model which is found in Shephard (1994). The O-SV model is described in more detail below.

Some of the notation in the O-SV model differs from that in the O-GARCH and O-EWMA models because not all of the previous notation is suited to the O-SV model. For example the conditional volatility of u_{it} is no longer represented as σ^2_{it} as it was in the two previous models but is given by $1/\gamma_{it}$ for $t = 1, 2, \cdots T$. Therefore in the O-SV model γ_{it} is the precision of u_{it} or, in the previous notation, γ_{it} is analogous to σ^{-2}_{it}. The model is described at times 0 and 1 and then extended to a general time t.

3.6.1 Model Description at times 0 and 1

The conditional distribution of the i^{th} precision at time 0, denoted γ_{i0}, is assumed to follow a gamma distribution with parameters $a_{i0} > 0$ and $b_{i0} > 0$

and is denoted as

$$\gamma_{i0}|F_0 \sim G(a_{i0},\ b_{i0}) \qquad \text{for } i = 1, 2, \cdots h.$$

The precision at time 1 relates to the precision at time 0 as follows

$$\gamma_{i1} = exp(r_{i1})\,\gamma_{i0}\,\eta_{i1} \tag{3.3}$$

with η_{i1} following a beta distribution such that

$$\eta_{i1} \sim beta(\omega_i\, a_{i0},\ (1-\omega_i)\, a_{i0}).$$

The notation used to represent the gamma and beta distributions above is explained further in Appendix A.1 along with the probability density functions. It is evident from the form of the beta distribution that the parameters of this distribution must be positive and therefore the constraint $0 < \omega_i < 1$ is required to ensure that both the beta parameters above are positive. The two distributions above also include the parameters a_{i0} and b_{i0} where each is the starting value of a time series of parameters. Lastly r_{i1} is a function of ω_i, a_{i0} and b_{i0}. This parameter r_{i1} is used to ensure that the mean of $log[\gamma_{i1}/\gamma_{i0}]$ is zero as discussed in detail at a later stage.

Updating equation (3.3) is used to move from the model at time 0 with distribution $\gamma_{i0}|F_0$ to the model at time 1 with distribution $\gamma_{i1}|F_1$. However, before determining the distribution of $\gamma_{i1}|F_1$ the distribution $\gamma_{i1}|F_0$ must be found. This can be done by applying statistical theory to the fact that γ_{i1} is the product of a non-random quantity $exp(r_{i1})$, a gamma random variable with parameters a_{i0} and b_{i0} and a beta random variable with parameters $\omega_i\, a_{i0}$ and $(1-\omega_i)\, a_{i0}$. Therefore the distribution of $\gamma_{i1}|F_0$ is

$$\gamma_{i1}|F_0 \sim G(\omega_i\, a_{i0},\ exp(-r_{i1})\, b_{i0}).$$

The proof is given in Appendix A.2.

Additionally there is one more distribution required in order to compute $\gamma_{i1}|F_1$. This is the distribution of $u_{i1}|\gamma_{i1}$ which is assumed to be

$$u_{i1}|\gamma_{i1} \sim N(0,\ \gamma_{i1}^{-1}).$$

Although this conditional distribution of the standardised principal component scores are normal, the information upon which the distribution is conditioned γ_{i1}^{-1} is different to that in the O-GARCH and O-EWMA models, that is F_0. Consequently this is a slight deviation from the general factor model specification in equation (2.7).

The distributions of $\gamma_{i1}|F_0$ and $u_{i1}|\gamma_{i1}$ has been introduced and now it is possible to calculate the distribution of $\gamma_{i1}|F_1$. Let $f(\cdot)$ denote a probability density function. Using the fact that $f(A|B) = f(A,B)/f(B)$, the probability density of $(\gamma_{i1}, F_1)|F_0$ can be broken up as

$$f(\gamma_{i1}, F_1|F_0) = f(\gamma_{i1}|F_0)f(F_1|\gamma_{i1}, F_0).$$

However $f(F_1|\gamma_{i1}, F_0)$ is equivalent to $f(u_{i1}|\gamma_{i1})$ since $F_0 \subset F_1$ and F_1 contains u_{i1} but F_0 does not. Therefore $f(\gamma_{i1}, F_1|F_0)$ can be determined because both $f(\gamma_{i1}|F_0)$ and $f(u_{i1}|\gamma_{i1})$ are known. Hence

$$\begin{aligned}
f(\gamma_{i1}, F_1|F_0) &= f(\gamma_{i1}|F_0)f(u_{i1}|\gamma_{i1}) \\
&= \left(\frac{\gamma_{i1}^{\omega_i a_{i0}-1}(exp(-r_{i1})b_{i0})^{\omega_i a_{i0}} exp(-\gamma_{i1} exp(-r_{i1})b_{i0})}{\Gamma(\omega_i\, a_{i0})}\right) \times \\
&\quad \left(\frac{1}{\sqrt{2\pi/\gamma_{i1}}} exp\left(-\frac{\gamma_{i1}}{2}u_{i1}^2\right)\right) \\
&= \left(\gamma_{i1}^{\omega_i a_{i0}-1} exp(-\gamma_{i1} exp(-r_{i1})b_{i0})\sqrt{\gamma_{i1}}\, exp\left(\frac{-u_{i1}^2\gamma_{i1}}{2}\right)\right) \times \\
&\quad \left(\frac{(exp(-r_{i1})b_{i0})^{\omega_i a_{i0}}}{\Gamma(\omega_i\, a_{i0})}\frac{1}{\sqrt{2\pi}}\right) \\
&= \left(\gamma_{i1}^{\omega_i a_{i0}+1/2-1} exp(-\gamma_{i1}[exp(-r_{i1})b_{i0} + \frac{1}{2}u_{i1}^2])\right) \times \\
&\quad \left(\frac{(exp(-r_{i1})b_{i0})^{\omega_i a_{i0}}}{\Gamma(\omega_i\, a_{i0})}\frac{1}{\sqrt{2\pi}}\right).
\end{aligned}$$

This shows that $f(\gamma_{i1}, F_1|F_0)$ can be written as the product of an expression containing γ_{i1} (first bracket) and an expression which does not containing γ_{i1} (second bracket). Bayes theorem implies that the distribution of $f(\gamma_{i1}|F_1)$ can be determined by looking at the mathematical form and parameters of the term containing γ_{it}. Hence the form of the first bracket is a gamma probability density function with parameters $(\omega_i\, a_{i0} + 1/2)$ and $(exp(-r_{i1})\, b_{i0} + \frac{1}{2}u_{i1}^2)$. Therefore the conditional distribution of γ_{i1} given F_1 is

$$\gamma_{i1}|F_1 \quad \sim \quad G(\omega_i\, a_{i0} + \tfrac{1}{2},\ exp(-r_{i1})\, b_{i0} + \tfrac{1}{2}\, u_{i1}^2).$$

Thus let $a_{i1} = \omega_i\, a_{i0} + \frac{1}{2}$ and $b_{i1} = exp(-r_{i1})\, b_{i0} + \frac{1}{2}u_{i1}^2$ to give

$$\gamma_{i1}|F_1 \quad \sim \quad G(a_{i1}, b_{i1}).$$

3.6.2 Model Description for a General Time t

To apply the steps above to a more general time t the conditional distribution of the standardised principal component scores, the updating equation and the distribution of η_{it} are required for a general time t. Firstly the conditional distribution of the factors is assumed to be

$$u_{it}|\gamma_{it} \sim N(0,\ \gamma_{it}^{-1})\,.$$

This conditional distribution does not totally conform to distribution (2.7) which is the conditional distribution that the standardised factors are assumed to follow in this thesis. This is also the case for time 1. The distribution does not conform to the general factor model as the conditional information is the precision γ_{it} and not F_{t-1}. However the O-SV model will still be considered to be a factor model even though it does not strictly meet the definition used in this thesis for a factor model.

Secondly the updating equation of the precision is needed along with the distribution of η_{it} for a general time t. The precision at time t relates to the precision at time $t-1$ as follows

$$\gamma_{it} = exp(r_{it})\,\gamma_{i,t-1}\,\eta_{it}$$

with the distribution of η_{it} assumed to be

$$\eta_{it} \sim beta(\omega_i\,a_{i,t-1},\ (1-\omega_i)\,a_{i,t-1}).$$

Now that the necessary information is available the same steps used to update $\gamma_{i0}|F_0$ to $\gamma_{i1}|F_1$ can be applied to the remaining times t. Applying these steps demonstrates that the distributions of $\gamma_{it}|F_t$ and $\gamma_{it}|F_{t-1}$ are both gamma distributions provided that the parameters a_{it} and b_{it} assume certain values (given below in equation (3.4)). Thus the conditional distributions of the precision are

$$\begin{aligned} \gamma_{it}|F_{t-1} &\sim G(\omega_i\,a_{i,t-1},\ exp(-r_{it})\,b_{i,t-1}) \qquad \text{and} \\ \gamma_{it}|F_t &\sim G(a_{it},\ b_{it}) \end{aligned}$$

provided that the parameters a_{it} and b_{it} are

$$\begin{aligned} a_{it} &= \omega_i\,a_{i,t-1} + \frac{1}{2} \qquad \text{and} \\ b_{it} &= exp(-r_{it})\,b_{i,t-1} + \frac{1}{2}\,u_{it}^2. \end{aligned} \tag{3.4}$$

Therefore the parameter values a_{it} and b_{it} ensure that the gamma densities are conjugate over time for each i, that is $\gamma_{it}|F_{t-1}$ and $\gamma_{it}|F_t$ each follow a gamma distribution for $i = 1, 2, \cdots h$ and all $t = 1, 2, \cdots T$. This conjugacy makes the model easier to work with. Additionally the parameters a_{it} and b_{it} can be represented as sums which are determined by repeated substitution in the recursive equations (3.4) to give

$$a_{it} = a_{i0}\,\omega_i^t + \frac{1}{2}\sum_{k=0}^{t-1}\omega_i^k \tag{3.5}$$

$$b_{it} = exp(-\sum_{k=1}^{t} r_{ik})\,b_{i0} + \frac{1}{2}\sum_{k=0}^{t-1}[(u_{i,t-k}^2)\prod_{l=0}^{k-1} exp(-r_{i,t-l})]. \tag{3.6}$$

Therefore the model can be summarised and represented in matrix form. For $i = 1, 2, \cdots, h$ and $t = 1, 2, \cdots, T$ the complete Gaussian local scale model of Shephard (1994) is

$$\boldsymbol{u_t^{(h)}}|\boldsymbol{\Gamma_t} \sim N(\boldsymbol{0}, \boldsymbol{\Gamma_t}^{-1})$$

$$\gamma_{it} = exp(r_{it})\,\gamma_{i,t-1}\,\eta_{it} \tag{3.7}$$

$$\eta_{it} \sim beta(\omega_i\,a_{i,t-1},\,(1-\omega_i)\,a_{i,t-1})$$

$$r_{it} = -E[\,log(\eta_{it})\,] = \psi(a_{i,t-1}) - \psi(\omega_i\,a_{i,t-1}) \tag{3.8}$$

$$\gamma_{i0}|F_0 \sim G(a_{i0},\,b_{i0})$$

where the ω_i's and the starting parameters a_{i0} and b_{i0} are unknown, $\psi(\cdot)$ is the digamma function and $\boldsymbol{\Gamma_t} = diag\{\gamma_{it}\}$. The parameter ω_i largely determines the rate at which the precision of the i^{th} standardised principal component score changes because the mean of η_{it} is ω_i for all t.

As mentioned for time 1, the rather curious parameter r_{it} is necessary to ensure that the conditional variance of the standardised principal component scores do not consistently display an upward or downward trend. This is desirable because in practice the conditional variance of the asset returns (and hence the standardised principal component scores) do not consistently exhibit an upward or downward trend. Therefore to ensure that the model has no such trends the expected log return of γ_{it} should be zero, i.e. $E[\,log(\gamma_{it}/\gamma_{i,t-1})\,] = 0$ for $t = 2, 3, \cdots, T$. The log ratio can be further decomposed using equation (3.7) to give $log(\gamma_{it}/\gamma_{i,t-1}) = r_{it} + log(\eta_{it})$. Therefore to ensure that the expected log returns of γ_{it} are zero $E[r_{it}]$ must equal $-E[\,log(\eta_{it})\,]$. If $r_{it} = -E[\,log(\eta_{it})\,]$ as in (3.8), then it will satisfy $E[r_{it}] = -E[\,log(\eta_{it})\,]$ as required. The value of $E[\,log(\eta_{it})\,]$ is calculated using the fact that if $Z \sim beta(\alpha,\, \beta)$ then $E[log(Z)] = \psi(\alpha) - \psi(\alpha + \beta)$.

3.6.3 Estimating the Conditional Variance

One of the major differences between a stochastic volatility model and a GARCH(1,1) or IGARCH(1,1) model is the form of the conditional variance. At each time point a stochastic volatility model outputs a distribution of the variance, in this case $\gamma_{it}|F_t$ or $\gamma_{it}|F_{t-1}$. On the other hand a GARCH(1,1) or IGARCH(1,1) model outputs the estimated conditional variance as a single value and not a distribution. Therefore when applying the O-SV model a value from the estimated conditional distribution of the precision needs to be calculated in order to obtain a single estimate of the conditional variance. For example the mean or median of the inverse of the precision are possible statistics which could be used to calculate a single point estimate of the conditional variance.

Consequently it may be considered intuitive to use $E[1/\gamma_{it}|F_{t-1}]$ or $E[1/\gamma_{it}|F_t]$ to estimate the conditional variance. However this does not tie in with the conditional variance used in the O-GARCH and O-EWMA models which is $\sigma^2_{it} = Var(u_{it}|F_{t-1})$. Therefore to estimate a comparable statistic in the O-SV model the distribution of $u_{it}|F_{t-1}$ is required. Shephard (1994) gives a method for calculating this distribution. This method is outlined below. The two key steps can be proved by applying Bayes Theorem. Thus

$$\begin{aligned} f(u_{it}|F_{t-1}) &= \int_0^\infty f(u_{it}, \gamma_{it}|F_{t-1})\, d\gamma_{it} \\ &= \int_0^\infty f(u_{it}|\gamma_{it}) f(\gamma_{it}|F_{t-1})\, d\gamma_{it} \end{aligned}$$

Since the distributions of $u_{it}|\gamma_{it}$ and $\gamma_{it}|F_{t-1}$ are both known, all that remains is to evaluate the integral to give

$$f(u_{it}|F_{t-1}) = \frac{1}{\sqrt{\frac{exp(-r_{it})b_{i,t-1}}{\omega_i a_{i,t-1}}}} \times \left[\frac{\Gamma\left(\frac{2\omega_i\, a_{i,t-1}+1}{2}\right)}{\Gamma\left(\frac{2\omega_i\, a_{i,t-1}}{2}\right)\sqrt{\pi\,(2\omega_i\, a_{i,t-1})}} \left(1 + \frac{u_{it}^2}{(2\,\omega_i a_{i,t-1})\,\frac{exp(-r_{it})b_{i,t-1}}{\omega_i a_{i,t-1}}}\right)^{-\frac{2\omega_i\, a_{i,t-1}+1}{2}} \right]. \quad (3.9)$$

Hence $u_{it}|F_{t-1}$ follows a central Student's t-distribution with $2\,\omega_i a_{i,t-1}$ degrees of freedom and scale parameter $exp(-r_{it})b_{i,t-1}/\omega_i a_{i,t-1}$. Once again let $\sigma_{it}^2 = Var[u_{it}|F_{t-1}]$, as this is the notation used in the previous two models. Therefore the conditional mean and variance of u_{it} are

$$\begin{aligned} E[u_{it}|F_{t-1}] &= 0 \\ \sigma_{it}^2 = Var[u_{it}|F_{t-1}] &= \left(\frac{exp(-r_{it})b_{i,t-1}}{\omega_i a_{i,t-1}}\right)\left(\frac{2\,\omega_i a_{i,t-1}}{2\,\omega_i a_{i,t-1} - 2}\right) \\ &= \frac{exp(-r_{it})b_{i,t-1}}{\omega_i a_{i,t-1} - 1}. \end{aligned} \quad (3.10)$$

However the conditional mean only exists if $\omega_i a_{i,t-1} > 0.5$ and the conditional variance only exists if $\omega_i a_{i,t-1} > 1$. For that reason a_{it} must be at least greater than one to calculate the conditional variance since ω_i lies between 0 and 1. This is a problem because the starting parameter a_{i0} is often close to zero, for reasons that are discussed in the next subsection. If $a_{i0} \approx 0$ then $a_{i1} = \omega_i\, a_{i0} + \frac{1}{2} \approx \frac{1}{2}$ since $0 < \omega_i < 1$ and hence the conditional variance (and possibly also the mean) do not exist at time 1 and possibly at some of the other early times too.

Although the conditional variance estimate (3.10) appears to be very different to the O-EWMA estimate, it is in fact fairly similar. This can be shown by using two approximations that will facilitate the interpretation of the conditional variance estimate. The first approximation is that if ω_i is relatively large then $exp(-r_{it})$ will be very close to ω_i (Shephard, 1994) and the second is to assume that $a_{i0} = 0$ and $b_{i0} = 0$. However these approximations are not actually used when implementing the model.

Firstly Shephard (1994) states that if ω_i is relatively large (say greater than

0.8) then $exp(-r_{it})$ will be very close to ω_i so that

$$\begin{aligned}\sigma^2_{it} &\approx \frac{\omega_i\, b_{i,t-1}}{\omega_i\, a_{i,t-1} - 1} \\ &= \frac{b_{i,t-1}}{a_{i,t-1} - 1/\omega_i}\,.\end{aligned}$$

Another implication is that under the assumption that ω_i is very close to 1, as time increases $a_{i,t-1}$ will become much larger than $1/\omega_i$ (which is approximately one) so that

$$\sigma^2_{it} \approx \frac{b_{i,t-1}}{a_{i,t-1}}\,.$$

By substituting $a_{i,t-1}$ with recursive equation (3.5) and $b_{i,t-1}$ with recursive equation (3.6) the estimate $\frac{b_{i,t-1}}{a_{i,t-1}}$ becomes

$$\frac{b_{i,t-1}}{a_{i,t-1}} = \frac{exp(-\sum_{k=1}^{t-1} r_{ik})\, b_{i0} + \frac{1}{2}\sum_{k=0}^{t-2}[(u^2_{i,t-k})\prod_{l=0}^{k-1} exp(-r_{i,t-l})]}{a_{i0}\,\omega_i^{t-1} + \frac{1}{2}\sum_{k=0}^{t-2}\omega_i^k}\,.$$

However this estimate can be further simplified by once again applying the approximation that if ω_i is relatively large then $exp(-r_{it})$ will be very close to ω_i (Shephard, 1994). Replacing $exp(-r_{it})$ with ω_i gives

$$\frac{b_{i,t-1}}{a_{i,t-1}} \approx \frac{b_{i0}\,\omega_i^{t-1} + \sum_{k=0}^{t-2}[(u^2_{i,t-k})\omega_i^k]}{a_{i0}\,\omega_i^{t-1} + \sum_{k=0}^{t-2}\omega_i^k}\,.$$

In addition if the approximation that $a_{i0} = 0$ and $b_{i0} = 0$ then

$$\frac{b_{i,t-1}}{a_{i,t-1}} \approx \frac{\sum_{k=0}^{t-2}[(u^2_{i,t-k})\omega_i^k]}{\sum_{k=0}^{t-2}\omega_i^k}$$

for $i = 1, 2, \cdots h$ and $t = 1, 2, \cdots T$.

Therefore under certain conditions the i^{th} conditional variance estimate (3.10) is approximately an exponentially weighted moving average of the squares of i^{th} series of standardised principal component scores. Hence the conditional variance estimates of the O-SV and O-EWMA models are analogous if ω_i is close to 1 and $a_{i0} \approx 0$ and $b_{i0} \approx 0$.

3.6.4 Parameter Estimation

The unknown parameters which need to be estimated are ω_i, a_{i0} and b_{i0} for all factors $i = 1, 2, \cdots h$. These parameters are estimated by maximising the log likelihood for each series separately. The log likelihood of the i^{th} series is

$$LL(u_{i1}, \cdots, u_{iT}, \omega_i, a_{i0}, b_{i0}) = -Tlog(\sqrt{2\pi}) + \sum_{t=1}^{t=T}[a_{it}log(exp(-r_{it})b_{i,t-1}/b_{it})$$
$$+log(\Gamma(a_{it})) - log(\Gamma(\omega_i a_{i,t-1})) - \frac{1}{2}log(exp(-r_{it})b_{i,t-1})]$$

which is found in Shephard (1994). There are a number of steps required to determine that this is in fact the likelihood.

Firstly the multivariate distribution of the i^{th} series of standardised principal component scores (i.e. the likelihood) is decomposed into the product of the conditional distributions $f(u_{it}|F_{t-1})$ using Bayes theorem. Hence the likelihood of the i^{th} series can be decomposed as

$$L(u_{i1}, \cdots, u_{iT}, \omega_i, a_{i0}, b_{i0}) = f(u_{iT}|F_{T-1})f(u_{i,T-1}|F_{T-2}) \cdots f(u_{i1}|F_0)f(F_0).$$

The second step is to find the conditional distributions $f(u_{it}|F_{t-1})$ and the distribution $f(F_0)$. The conditional distributions have already been calculated in the previous subsection and the distribution $f(F_0)$ is assumed to be 1. Taking the log of $L(u_{i1}, \cdots, u_{iT}, \omega_i, a_{i0}, b_{i0})$ gives the log likelihood, as required.

Although this calculation gives an analytical formula for the log likelihood of the i^{th} series, it cannot be maximised analytically with respect to the three unknown parameters ω_i, a_{i0} and b_{i0}. Therefore the values of ω_i, a_{i0} and b_{i0} which maximise the likelihood must be found using numeric techniques. However using numerical techniques to maximise the log likelihood with respect to all three parameters simultaneously, for a given series, does not result in the log likelihood converging to a maximum. This occurred for the data in this thesis when using the MATLAB optimiser, although this may not be the case in general. The multivariate optimisation routine (in MATLAB) simply diverges.

For this reason the parameters a_{i0} and b_{i0} are assigned values. This is of little consequence because the value of the log likelihood appears to be fairly insensitive to the starting parameters a_{i0} and b_{i0}. This is demonstrated in the results section. In addition the choice of a_{i0} and b_{i0} do not make a big difference to the model parameters a_{it} and b_{it}, except for the first few time steps. This is because as time t gets larger so the first term of a_{it} (see (3.5)) and b_{it} (see (3.6)) tend to zero since ω_i^t will tend to 0 as t increases because $0 < \omega_i < 1$. Moreover a similar approach was used for the starting value σ_{i1}^2 in the GARCH(1,1) and the IGARCH(1,1) models, that is a value is assigned to the starting value σ_{i1}^2 instead of estimating it by maximising the likelihood.

One possible choice for the starting values are $a_{i0} = 0$ and $b_{i0} = 0$, which can be justified in the Bayesian paradigm because these values would result in a non-informative prior distribution that $\gamma_{i0}|F_0$ is a non-informative prior. Shephard (1994) does not explicitly mention which values he assigns to a_{i0} and b_{i0} but merely states that Harvey (1989) assigns $a_{i0} = 0$ and $b_{i0} = 0$ in a related model setting. However, choosing $a_{i0} = 0$ and $b_{i0} = 0$ exactly results in problems estimating r_{i1} as a minus infinity, plus infinity situation arises. This is evident in equation (3.8). Therefore a value close to zero but not exactly zero is used for both.

3.6.5 Forecasting the Conditional Volatility

Once the model has been fitted to obtain parameter estimates the conditional volatility can be forecasted using the information up to that point in time. Let $\sigma_{i,t+k|t}^2$ be the forecast of the conditional variance at time $t+k$ given the information available at time t. The one step ahead forecast can be derived using the probability density function $f(u_{i,t+1}|F_t)$ which has already been derived as the density in equation (3.9). Therefore in the case of $k = 1$

$$\sigma_{i,t+1|t}^2 \quad = \quad \frac{exp(-r_{i,t+1})b_{it}}{\omega_i a_{it} - 1}\,.$$

On the other hand, to determine $\sigma_{i,t+k|t}^2$ for $k > 1$ the distribution of $f(u_{i,t+k}|F_t)$ is required but to calculate this the distribution of $f(\gamma_{i,t+k}|F_t)$ is needed. However the distribution of $f(\gamma_{i,t+k}|F_t)$ for $k > 1$ cannot be found without knowledge of future observations of the standardised principal component scores $u_{i,t+l}$ for $l > 0$. This is because these observations are inherent in $b_{i,t+l}$ for $l > 0$.

Since neither of the distributions $f(u_{i,t+k}|F_t)$ or $f(\gamma_{i,t+k}|F_t)$ can be determined an alternative method to calculate $\sigma_{i,t+k|t}^2$ needs to found. One possible solution is to use $1/E(\gamma_{i,t+k} \,|\, \gamma_{it})$ as an estimate of $\sigma^2_{i,t+k|t}$ since the distribution $f(\gamma_{i,t+k} \,|\, \gamma_{it})$ can be determined. The reason that $E(\gamma_{i,t+k} \,|\, \gamma_{it})$ is used instead of $E(\frac{1}{\gamma_{i,t+k}} \,|\, \gamma_{it})$ is because the latter requires a certain criterion to be satisfied but this criterion is not always met. This is discussed in more detail below.

However, there are two difficulties with using $E(\gamma_{i,t+k} \,|\, \gamma_{it})$ as an estimate for $\sigma^2_{i,t+k|t}$. The first is that this forecast will not be directly comparable to the O-EWMA and O-GARCH forecasts as these are $\sigma^2_{i,t+k|t} = Var(u_{i,t+k}|F_t)$ and not $\sigma^2_{i,t+k|t} = 1/E(\gamma_{i,t+k} \,|\, \gamma_{it})$ as is the suggested case for the O-SV model. The second is that the formula involves γ_{it} which is assumed to be known, when in fact only an estimate of γ_{it} exists. Therefore $E[\gamma_{it}|F_t]$ is used as a point estimate of γ_{it} in the formula in place of the actual value as it is unknown.

There are two ways to calculate $E(\gamma_{i,t+k} \,|\, \gamma_{it})$ (Shephard, 1994). The first method uses the standard stochastic volatility model which has already been discussed. However the second method involves adapting the standard stochastic volatility model to deal with missing observations or irregularly spaced data.

Method 1
To calculate $E(\gamma_{i,t+k} \,|\, \gamma_{it})$, $\gamma_{i,t+k}$ needs to be represented as a function of γ_{it} only and none of the other $\gamma_{i,t+l}$'s for $t < l < k$. Such an equation can be found by repeated substitution into the updating equation (3.7) to give

$$\gamma_{i,t+k} \;\; = \;\; \gamma_{it} \left[\prod_{l=1}^{l=k} exp(r_{i,t+l}) \, \eta_{i,t+l} \right].$$

This equation contains the variables $\eta_{i,t+l}$ which each follow a beta distribution with parameters $\omega_i \, a_{i,t+l-1}$ and $(1 - \omega_i) \, a_{i,t+l-1}$ for $1 \leq l \leq k$. In addition the equation contains $r_{i,t+l}$ for $1 \leq l \leq k$ where each $r_{i,t+l}$ is also a function of the parameters ω_i and $a_{i,t+l-1}$. Therefore both $\eta_{i,t+l}$ and $r_{i,t+l}$ are only a function of the a's and ω_i's which do not depend on future observations so there are no problems with unknown future observations. Hence $E(\gamma_{i,t+k} \,|\, \gamma_{it})$ can be calculated using properties of the beta distribution to give

$$E[\gamma_{i,t+k} \,|\, \gamma_{it}] \;\; = \;\; \gamma_{it} \left[\prod_{l=1}^{l=k} exp(r_{i,t+l}) \right] \, \omega_i^k \, .$$

However this assumes that γ_{it} is known when in fact it is not. Hence $E[\gamma_{it}|F_t] = a_{it}/b_{it}$ is used as a point estimate of γ_{it} so that

$$E[\gamma_{i,t+k} \,|\, \gamma_{it}] \;\approx\; \frac{a_{it}}{b_{it}} \left[\prod_{l=1}^{l=k} exp(r_{i,t+l}) \right] \omega_i^k.$$

The reason that $E(1/\gamma_{i,t+k} \,|\, \gamma_{it})$ is not used as an estimate of $\sigma^2_{i,t+k|t}$ is because it requires that $\omega_i a_{i,t+l} > 1$ for $l = 0, \cdots, k-1$. This translates to requiring $a_{i,t+l}$ to be at least greater than 1 if not larger for $l = 0, \cdots, k-1$ which is a problem because this condition is not always met.

Method 2

In contrast the second method does not use the standard stochastic volatility model but adapts it to allow for missing or irregularly spaced observations. One way to think about it is that there are observations at times $1, 2, \cdots, t$ and then the next observation is only at time $t+k$ so that the data are essentially irregularly spaced. In this adjusted model let time τ be a symbol which actually represent time t_τ so that the difference between time τ and $\tau+1$ is not necessarily one time step but may be more than one time step. The details of the adjusted model with irregularly spaced/missing observations are given below for $i = 1, 2, \cdots, h$ (Shephard, 1994).

$$\begin{aligned}
\boldsymbol{u_\tau^{(h)}}|\boldsymbol{\Gamma_\tau} &\sim N(\mathbf{0}, \boldsymbol{\Gamma_\tau}^{-1}) \\
\gamma_{i\tau} &= exp(r^*_{i\tau})\,\gamma_{i,\tau-1}\,\eta_{i\tau} \\
\Delta_\tau &= t_\tau - t_{\tau-1} \\
\eta_{i\tau} &\sim beta(\omega_i^{\Delta_\tau}\, a^*_{i,\tau-1},\, (1-\omega_i^{\Delta_\tau})\, a^*_{i,\tau-1}) \\
r^*_{i\tau} = -E[\,log(\eta_{i\tau})\,] &= \psi(a^*_{i,\tau-1}) - \psi(\omega_i^{\Delta_\tau}\, a^*_{i,\tau-1}) \\
\gamma_{i0}|F_0 &\sim G(a^*_{i0},\, b^*_{i0})
\end{aligned}$$

with $a^*_{i\tau} = \omega_i^{\Delta_\tau}\, a^*_{i,\tau-1} + \frac{1}{2}$ and $b^*_{i\tau} = exp(-r^*_{i\tau})\, b^*_{i,\tau-1} + \frac{1}{2}u^2_{i\tau}$. The stars * indicate that the parameters are from the adjusted stochastic volatility model with irregularly spaced data and not from the standard stochastic volatility model with regular spaced data. The parameter $r^*_{i\tau}$ is calculated using a different formula to the standard model but with the same purpose of ensuring that the expected log return of the precision is zero. It is calculated using the fact that if $Z \sim beta(\alpha, \beta)$ then $E[log(Z)] = \psi(\alpha) - \psi(\alpha + \beta)$.

Therefore using the information in the adjusted model above $E[\gamma_{i,\tau+1} \,|\, \gamma_{i\tau}]$ can be calculated to give

$$E[\gamma_{i,\tau+1} \,|\, \gamma_{i\tau}] \;\; = \;\; \gamma_{i\tau} \, exp(r^*_{i,\tau+1}) \, \omega_i^{\Delta_{\tau+1}}.$$

As previously mentioned, the expectation assumes γ_{it} is known when it is not. Therefore $\gamma_{i\tau}$ is again replaced by the single point estimate $E[\gamma_{i\tau}|F_\tau] = a^*_{i\tau}/b^*_{i\tau}$. If this is substituted into $E[\gamma_{i,\tau+1} \,|\, \gamma_{i\tau}]$ then this expectation becomes

$$E[\gamma_{i,\tau+1} \,|\, \gamma_{i\tau}] \;\; \approx \;\; \frac{a^*_{i\tau}}{b^*_{i\tau}} \; exp(r^*_{i,\tau+1}) \, \omega_i^{\Delta_{\tau+1}}.$$

Hence if $t_\tau = t$ and $t_{\tau+1} = t + k$ then $\Delta_{\tau+1} = k$ so this result reduces to

$$E[\gamma_{i,\tau+1} \,|\, \gamma_{i\tau}] \;\; \approx \;\; \frac{a^*_{i\tau}}{b^*_{i\tau}} \; exp(r^*_{i,\tau+1}) \, \omega_i^{k}.$$

Once again the reason that $E[1/\gamma_{i,t+k} \,|\, \gamma_{it}]$ is not used as an estimate of $\sigma^2_{i,t+k|t}$ is because it requires that $\omega_i^{\Delta_\tau} \, a^*_{i,\tau-1} > 1$. This translates to the requirement that $a^*_{i,\tau-1}$ is at least greater than 1 or possibly even larger which is a problem because this condition is not always met.

It is important to note that $t_1 = 1, t_2 = 2, \cdots, t_\tau = t$ because there are no irregularly spaced data between time 1 and time t. Therefore $a^*_{i\tau} = a_{it}$ and $b^*_{i\tau} = b_{it}$ because $\Delta_\tau = 1$ for $t_\tau = 1, 2, \cdots, t$. In this case the estimate of $E[\gamma_{i,\tau+1} \,|\, \gamma_{i\tau}]$ with $t_\tau = t$ and $t_{\tau+1} = t + k$ becomes

$$E[\gamma_{i,\tau+1} \,|\, \gamma_{i\tau}] \;\; \approx \;\; \frac{a_{it}}{b_{it}} \; exp(r^*_{i,\tau+1}) \, \omega_i^{k}.$$

This result can be compared to the result in the first method where there are no missing observations which is

$$E[\gamma_{i,t+k} \,|\, \gamma_{it}] \;\; \approx \;\; \frac{a_{it}}{b_{it}} \left[\prod_{l=1}^{l=k} exp(r_{i,t+l}) \right] \omega_i^{k} \, .$$

Hence the only difference between the results from the two methods is that the one estimate of $E[\gamma_{i,t+k} \,|\, \gamma_{it}]$ contains $\left[\prod_{l=1}^{l=k} exp(r_{i,t+l})\right]$ and in place of this the other contains $exp(r^*_{i,\tau+1})$. However, Shephard (1994) shows that these two estimates are similar for ω_i around 0.93 provided that the forecast is not too far ahead, that is k is not too large.

In this thesis the second method, which assumes irregularly spaced data, is used to estimate $E[\gamma_{i,t+k} \,|\, \gamma_{it}]$ so that $\sigma^2_{i,t+k|t}$ is estimated using $1/E[\gamma_{i,t+k} \,|\, \gamma_{it}]$.

To summarise, the one step ahead variance $\sigma^2_{i,t+1|t}$ is forecasted using $1/E[\gamma_{i,t+1}|F_t]$ and the k step ahead forecast $\sigma^2_{i,t+k|t}$ where $k > 1$ is estimated using $1/E[\gamma_{i,\tau+1} \,|\, \gamma_{i\tau}]$ where $t_\tau = t$, $t_{\tau+1} = t + k$ and all observations before time t have a time step of one between them. However for the case $k = 1$ the estimate of $1/E[\gamma_{i,t+1}|F_t]$ is the same as the estimate for $1/E[\gamma_{i,\tau+1} \,|\, \gamma_{i\tau}]$. Therefore for all integers $k > 0$ the forecasted conditional variance is

$$\sigma^2_{i,t+k|t} \;=\; 1/E[\gamma_{i,t+k} \,|\, \gamma_{it}] \approx \frac{b_{it}}{a_{it}\; exp(r^*_{i,\tau+1})\,\omega_i^k}\,.$$

Chapter 4

Details of the Model Adjustments

This chapter discusses adjustments made to the raw data as well as adjustments made at various stages in the calculations for fitting the orthogonal factor models. There are three main steps in the calculations in which adjustments can be made. These three stages are the calculation of the vectors $\boldsymbol{z_t}$, $\boldsymbol{y_t}$ and $\boldsymbol{u_t}$, following the notation introduced in Figure (2.1)). These three stages and the associated vectors are:

1. The method used to calculate the log returns $\boldsymbol{z_t}$.

2. The method used to adjust the log returns $\boldsymbol{z_t}$ to obtain the adjusted residuals $\boldsymbol{y_t}$.

3. The method used to adjusts the factors $\boldsymbol{x_t}$ to obtain the standardised factors $\boldsymbol{u_t}$.

4.1 Step 1: Return Calculations

The orthogonal factor models are fitted to two data sets, the first data set contains daily closing share prices and the second daily exchanges rates. Both of these are expected to display some sort of trend over time. Share prices typically increase over time and exchange rates can increase or decrease over time depending on the relative inflation and interest rates of the two regions, among other things. Thus the data as is, are not conducive to statistical analysis. To overcome this problem the data are converted to continuous log returns which should not display a trend over time. The log return calculations are discussed in this section.

Thus let P_t be the closing share price or exchange rate on day t. Therefore the annualised log return Ret_t at time t is calculated as

$$Ret_t = 250 * [ln(P_t) - ln(P_{t-1})].$$

The 250 is there to convert the returns from continuous daily returns to continuous annualised returns based on the assumption that on average there are 250 trading days in a year. The returns are annualised because it is more intuitive to work with annualised returns.

However in the case of shares, the return calculations are a bit more complicated than this when there are dividends or share splits.

4.1.1 Adjustment to Returns for Dividends

The dividend adjustment discussed here is for a cash dividend as the value of the dividend is known on the declaration date. This is important because the monetary value of dividends paid in another form may not be known on the declaration date, for example dividends paid in the form of shares. The reason an adjustment is required is discussed before describing the adjustment itself.

On the last day to trade (LDT) an investor who purchases a share will receive a dividend D. However if an investor purchases the share the next day then the investor will not receive the dividend. Therefore purchasing the share on the LDT involves purchasing a certain percentage of the company plus the promise of cash D at a known future date. However purchasing a share the following day only involves purchasing that same percentage of the company but without the cash payment D. Hence these two share purchases are not comparable due to the difference of the future cash payment D. As a result the log return needs to be calculated in a manner which allows for the dividend so that the purchases are comparable.

Thus some theory is considered before a practical solution is given. To develop the theory it is assumed that stock markets do not close. Let time t be the end of the t^{th} day which is the LDT for the cash dividend of amount D. Therefore up to and including time t the purchaser will receive the dividend but not after time t. Hence one may think that

$$P_t^- = P_t^+ + \text{Present Value of } D \tag{4.1}$$

where P_t^- is the share price at the instant before the time t and P_t^+ is the share price the instant after time t. If this statement were true then the daily log return from time t^- to t^+ is

$$log(P_t^+) - log(P_t^-)$$

which is not zero if equation (4.1) is true. Consequently, assuming equation (4.1) is true, the return is only non zero because of the dividend and not because of any other change in the underlying company. Thus if an adjustment is not made for the dividend then the return series may appear more volatile than it should be. Additionally these two purchases are not comparable but this calculation implicitly assumes that they are. Hence if equation (4.1) is true then the return from time t^- to time t^+ should be zero. Therefore it should be calculated as

$$log(P_t^+ + \text{Present Value of } D) - log(P_t^-)$$

to ensure that the dividend does not incorrectly affect the return.

However these calculations above assume that equation (4.1) is true but it is not the case since it assumes that the share price P_t^+ is known at time t^-. This is incorrect as the share price is not previsible so P_t^+ cannot be known at time t^-. One possible way to overcome this is to use forward prices instead of share prices because forward prices are previsible. If the P_t's are forward prices then formula (4.1) is valid. The reason forward prices are not used to calculate returns is because they depend on the strike price and maturity date and not just on the underlying share price. Therefore there is a conflict between practical requirements and theoretical accuracy when calculating the returns.

For practical purposes equation (4.1) is assumed to be true. However even if equation (4.1) is true, it cannot be applied as is. This is because in reality markets do close and finding an appropriate rate to use for the purposes of discounting the dividend is tricky and the rate may be inaccurate. Hence equation (4.1) needs to be tweaked for practical application.

Firstly in practice stock markets do close so P_t, the closing share price on day t, replaces P_t^- and P_{t+1}, the closing price on day $t+1$, is used to replace P_t^+. P_{t+1} is used to replace P_t^+ instead of the next available price which would be the opening price on day $t+1$ because daily returns are required.

Secondly calculating the present value of each dividend is not practical. One reason is that it is very time intensive to discount each dividend because the interest rate of an appropriate term at that time which reflects a similar level of risk to the company must be determined. In fact it may not be possible to calculate such a rate which meets all the necessary criteria and is calculated in an objective manner. Additionally the difference between the present value of the dividend and the actual value D is very small because the time between the LDT and the payment date is usually short (a week or two). Therefore whether the dividend is discounted or not should not have a material impact. Hence in practice the value of the dividend D is used instead of the present value.

Thus if day t is the LDT for the dividend of amount D then for practical purposes the log return is calculated as

$$Ret_{t+1} \quad = \quad 250 * [ln(P_{t+1} + D) - ln(P_t)]. \tag{4.2}$$

Although this is not theoretically correct, for practical purposes adjustment (4.2) is reasonable and probably close to the best which can be achieved given the practical difficulties.

4.1.2 Adjustment to Returns for Share Splits

In addition to these difficulties share splits also need to be considered. In the case of share splits one share after the split will no longer reflect the same proportion of the company as one share before the split therefore an adjusted price needs to be determined. If there is an m for n share split at the start of day t then for every n shares the owner held at the end of day $t-1$ the owner will have m at the start of day t. Hence the return should be calculated as

$$Ret_t \quad = \quad 250 * [ln(P_t * \frac{m}{n}) - ln(P_{t-1})]$$

where P_t is the published closing share price at time t which reflects the price after the split.

4.2 Step 2: Calculation of the Adjusted Residuals $\boldsymbol{y_t}$

The calculation and in particular the adjustment of the returns $\boldsymbol{z_t}$ have been dealt with so the next step is to calculate the residuals $\boldsymbol{y_t}$ using the returns

$\boldsymbol{z_t}$. Recall in Figure (2.1) that $\boldsymbol{z_t}$ and the adjusted residuals $\boldsymbol{y_t}$ are related as

$$\boldsymbol{y_t} \;=\; \boldsymbol{B}\left(\boldsymbol{z_t} - \boldsymbol{\hat{\mu}_t}\right)$$

where $\boldsymbol{\hat{\mu}_t}$ is the estimated conditional mean of $\boldsymbol{z_t}$ and $\boldsymbol{B}$ is an $N \times N$ invertible matrix used to standardise the residuals. Thus the first adjustment to $\boldsymbol{z_t}$ involves removing the conditional mean and the second involves taking a linear combination of the residuals.

4.2.1 Removing the Conditional Mean

As previously discussed in Section 2.2, the conditional mean of the returns $\boldsymbol{z_t}$ should in theory be zero if markets are arbitrage free. However because this is not always the case in practice, the conditional mean of the returns $\boldsymbol{z_t}$ needs to be removed in the process of calculating $\boldsymbol{y_t}$. This is because model (2.5) assumes that the factors $\boldsymbol{x_t}$ have a zero conditional mean which requires that the adjusted residuals $\boldsymbol{y_t}$ need to have a zero conditional mean.

The conditional mean is removed from the data by fitting an ARMA(p,q) model or, as suggested by Fan et al. (2008), a VAR(p) model to the returns. For each of these two models fitted (ARMA(p,q) and VAR(p)) a method for choosing the values of p and of q in the former are required. In the case of the ARMA(p,q) model the values of p and q are chosen separately for each series.

A quick and dirty method is used to choose the value of p and where applicable q. It involves using the Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) and the Likelihood Ratio Test. The values of p and q are chosen by selecting the model with either the smallest Akaike Information Criteria (AIC) or the smallest Bayesian Information Criteria (BIC) up to and including p and q equal to six. However for sample sizes greater than 7 the BIC imposes a greater penalty than the AIC does for additional model parameters. This is applicable in this thesis as the sample size of each of the data sets is greater than 7.

Therefore in the case of a VAR(p) model, selecting a model using the BIC will result in the same or a smaller value for p than that selected by the AIC. If both criteria indicate using the same value of p then that value is used but if the one suggested by the BIC is smaller then a likelihood ratio test is performed to determine whether the additional parameter or parameters are zero at the 5% significance level. If this test indicates that the parameter or parameters are significantly different from zero then the value of p as recommended by the AIC is used, otherwise the value of p as recommended by the BIC is used.

Similarly in the case of the ARMA(p,q) model, selecting a model using the BIC will usually result in the same or a smaller values for p and q. However in this case there are two parameters so the model selected by the BIC will not necessarily be nested in the one selected by the AIC. Therefore if the model selected by the BIC is not nested in the model selected by the AIC then it is not possible to perform a likelihood ratio test on the two models selected. In such a situation the values of p and q recommended by the BIC are used. On the other hand if the model selected by the BIC is nested in the model selected by the AIC then a likelihood ratio test is performed to determine whether the additional parameter or parameters are zero at the 5% significance level. If this test is significant then the values of p and q recommended by the AIC are used, otherwise the values of p and q recommended by the BIC are used.

Hence the model selection procedure for both the ARMA(p,q) and the VAR(p) models are very similar. Although these model selection procedures are unacceptable if the focus of the thesis is only on fitting these models, they are adequate for the purpose required. In other words these models are simply used to remove the conditional mean which in theory should be zero anyway. Hence in the bigger picture using this method as opposed to a more rigorous model selection process should have little impact on the fit of the conditional covariances.

4.2.2 Linear Combination of the Residuals

Executing the steps above will provide an estimate of the sample residuals

$$\hat{\boldsymbol{e}}_{\boldsymbol{t}} = \boldsymbol{z}_{\boldsymbol{t}} - \hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$$

where $\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$ is an estimate of $E[\boldsymbol{z}_{\boldsymbol{t}}|F_{t-1}]$. Instead of converting theses residuals directly to factors, it may be preferable to take a linear combination of these for reasons that will be discussed shortly. Therefore the adjusted residuals $\boldsymbol{y}_{\boldsymbol{t}}$ are calculated by multiplying the residuals by a chosen value of the matrix $\boldsymbol{B}$ (see equation (2.3)).

In this thesis two different values for $\boldsymbol{B}$ are tested. These two along with the resulting implications for the adjusted residuals $\boldsymbol{y_t}$ are described.

1. Each residual $\hat{e_{it}}$ is divided by the sample standard deviation of that series of residuals (i.e. the i^{th} series), as suggested by Alexander (2000). Therefore each series of the adjusted residuals $\boldsymbol{y_t}$ has a zero sample mean (since the mean of the residuals is zero) and a sample variance of one. For this reason these adjusted residuals $\boldsymbol{y_t}$ are called the standardised residuals. Recall that $\hat{\boldsymbol{\Phi}}$ is the unconditional sample covariance matrix of the sample residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$, as introduced in Section 2.3. Hence let $\hat{\phi}_{ij}$ be the sample covariance of the i^{th} and j^{th} series of residuals. Therefore to construct the standardised residuals $\boldsymbol{y_t}$ the matrix $\boldsymbol{B} = diag\left\{\hat{\phi}_{ii}^{-\frac{1}{2}}\right\}$ is invoked.

2. The residuals are not adjusted so $\boldsymbol{B}$ is the identity matrix. If this choice is made then $\boldsymbol{y_t}$ is simply the unadjusted residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$.

There is a possible third choice for $\boldsymbol{B}$ which could not be used because of the difficulties which result when calculating the principal components. This involves adjusting the residuals using the sample covariance matrix of the residuals, as suggested by Fan et al. (2008). Thus $\boldsymbol{B} = \hat{\boldsymbol{\Phi}}^{-\frac{1}{2}}$ so that $\boldsymbol{y_t}$ will be the product of $\hat{\boldsymbol{\Phi}}^{-\frac{1}{2}}$ and the sample residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$. This results in the sample covariance matrix of $\boldsymbol{y_t}$ being the identity matrix. The implication is that all the eigenvalues are one and any vector of the appropriate dimensions with unit length will be an eigenvector. Therefore the principal components will consist of any set of orthogonal vectors of the appropriate dimension with unit length. van der Weide (2002) suggests that in such cases identification problems occur with $\boldsymbol{A}$.

Some of the reasoning behind the remaining two alternatives as introduced above are considered. The first option can be used since it was suggested by Alexander (2000). Moreover before computing the principal components of any data set it is typical for each series to be standardised to have a sample mean of zero and a sample variance of one (Manly, 2005). The reason for this is to ensure that the variance does not have an unwarranted influence on the principal components. Hence the eigenvalues and eigenvectors will be those of the sample correlation matrix instead of the sample covariance matrix, as is the case with the second choice.

In some circumstances it is believed that the variance of a series reflects the importance of that series (Manly, 2005). Under these circumstances the variances should not be standardised to one. Although the variance may not reflect the importance of a series in the case of the data used in this thesis, the variance of a series may well reflect important information and this is one reason for the second choice above. The other reason for the second choice is simplicity as no additional calculations are necessary and an adjustment may not be necessary.

4.3 Step 3: Adjustment of the factors $\boldsymbol{x_t}$

Once one of the possible alternatives for $\boldsymbol{B}$ has been selected and consequently the adjusted residuals vector $\boldsymbol{y_t}$ has been calculated, the factors $\boldsymbol{x_t}$ can be constructed as the principal component scores of $\boldsymbol{y_t}$. However before modelling the conditional variances one final adjustment is required. This is to calculate the standardised factors $\boldsymbol{u_t}$. These standardised factors are constructed as a linear combination of the factors $\boldsymbol{x_t}$ using matrix $\boldsymbol{W}$ as introduced in Figure (2.1)).

There are two possible choices used for $\boldsymbol{W}$ in this thesis. The first is to let $\boldsymbol{W} = \boldsymbol{I}$ in which case the factors are actually not adjusted. The second is to standardise the factors so that each series of factors has unit sample variance. This adjustment was suggested by Bongers (2008) in the context of an O-GARCH model.

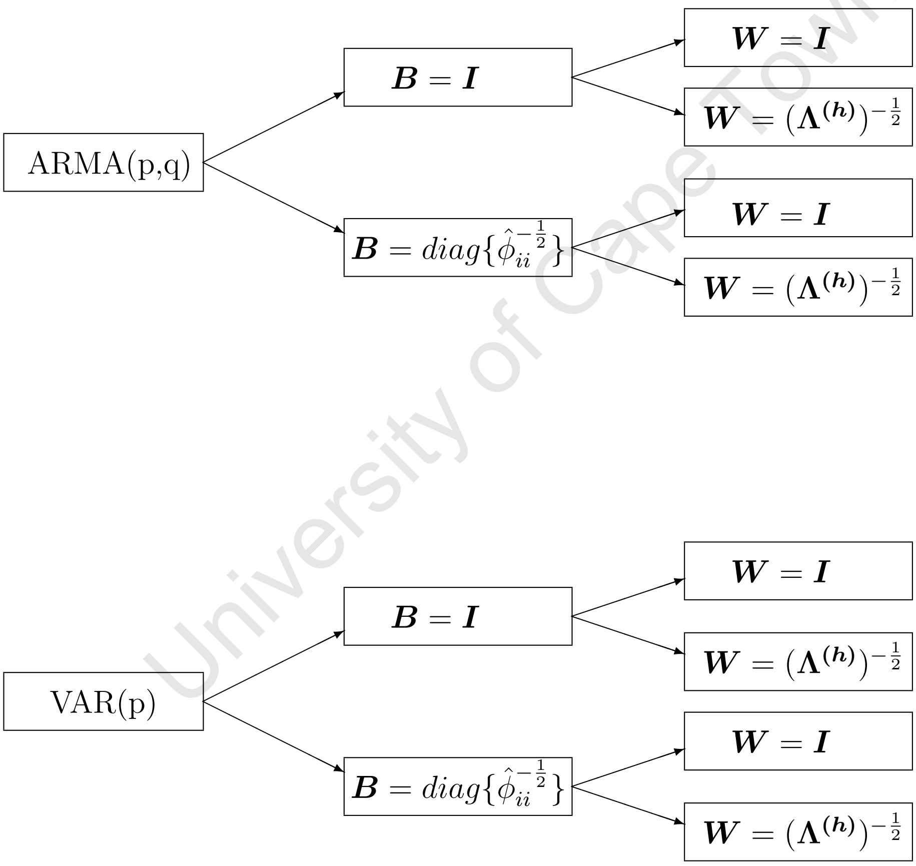
A brief outline of the reasons for this adjustment are discussed. Bongers (2008) uses simulation to demonstrate that the square parameter error of a GARCH(1,1) model is smaller on average when this adjustment is made. This suggests that the adjustment will improve the fit of the O-GARCH model. However Bongers (2008) did not test whether this adjustment could improve the fit of the O-EWMA or O-SV models. Hence the fit of these two models using this adjustment will be tested on the two data sets introduced in the present study, although no conclusions can be drawn for a general data set from such tests. This adjustment is considered in more detail as follows. Recall that it was demonstrated in Section 3.1 that the unconditional sample variance of the i^{th} series of principal component scores is the i^{th} eigenvalue λ_i. Hence each series of principal component scores are simply divided by their sample standard deviation. To represented this adjustment mathematically let $\boldsymbol{\Lambda}^{(\boldsymbol{h})} = diag\{\lambda_i\}$, where λ_i is the i^{th} eigenvalue of $\hat{\boldsymbol{V}}$ such that $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_h$ for $i = 1, 2, \cdots, h$. Therefore let $\boldsymbol{W} = (\boldsymbol{\Lambda}^{(\boldsymbol{h})})^{-\frac{1}{2}}$ such that the principal component scores $\boldsymbol{x_t}$ are multiplied by $(\boldsymbol{\Lambda^{(h)}})^{-\frac{1}{2}}$ to give

$$\boldsymbol{u_t} \quad = \quad (\boldsymbol{\Lambda^{(h)}})^{-\frac{1}{2}}\boldsymbol{x_t}\,.$$

4.4 Summary of All the Possible Adjustments

For each of the three models, O-GARCH, O-EWMA and O-SV, there are 8 different model variations. These are given in the flow chart depicted in Figure 4.1. Hence there are 8 different estimates of the conditional covariance of the log returns $\boldsymbol{z_t}$ at each time t. Therefore in total there are 24 different outcomes to compare for each of the data sets. This clearly makes the comparison of results difficult. For this reason it is not possible to graphically compare all of the conditional covariance estimate for all the models. However the statistics are fitted to all of the model variations considered.

Figure 4.1: Chart of All Possible Model Variations



Chapter 5

Overview of the Data Sets

Two data sets each consisting of ten years of daily data from the 21st May 1999 to 20th May 2009 have been used. A description of the exact data contained within each data set as well as their characteristics are considered in this chapter.

The purpose of using two data sets is that this allows two different types of financial data to be modelled which gives an indication of how the model fit varies for different types of data. Hence if a model fits both types of data well then it may be considered preferable to a model which only fits one type of data well. The reasoning behind this is that in practice a model should fit a variety of financial data well because the required conditional covariance matrix will typically be that for a variety of financial data and not just one type. This is because the practitioner is usually interested in the relationship between various financial data. If the model only fits one type of data well then this would make it difficult to construct a reasonable covariance matrix of different asset classes and other financial data.

5.1 General Description of the Share Data

The first data set consists of the daily closing prices of seven shares which are listed on the Johannesburg Stock Exchange (JSE) from the 21st May 1999 to 20th May 2009. These data were obtained from McGregor BFA. The names of these shares along with the sector in which they are classified are given in Table 5.1.

It should be noted that four of the seven shares are from the banking sector.

Table 5.1: Data Set 1 - Shares Selected and their Sectors

Share	**JSE Sector**
ABSA Group Limited	Banks
FirstRand Limited	Banks
Standard Bank Group Limited	Banks
Nedbank Group Limited	Banks
Gold Fields Limited	Gold Mining
Murray and Roberts Holdings Limited	Other Construction
Pick 'n Pay Stores Limited	Food and Drug Retailer

Initially the idea was that this would allow the share data set to be used as is and also to be split into two groups. However there are already 24 model estimates (see Section 4.4) which need to be tested for each data set and splitting this data set in addition to using it as is will create an additional 48 model estimates which need to be tested. For this reason the data set is not split into two in this thesis. However this is a possible extension which could facilitate further testing of the models.

Table 5.2: Correlation of the Share Returns

	ABSA	**First**	**StdB**	**Nedb**	**GFlds**	**M&R**	**P'nP**
ABSA	1.00	0.59	0.58	0.54	-0.02	0.23	0.28
FirstRand	0.59	1.00	0.65	0.57	0.02	0.28	0.32
Std Bank	0.58	0.65	1.00	0.58	-0.01	0.25	0.30
Nedbank	0.54	0.57	0.58	1.00	0.00	0.25	0.29
Gold Fields	-0.02	0.02	-0.01	0.00	1.00	0.05	0.06
Murray&Rob	0.23	0.28	0.25	0.25	0.05	1.00	0.17
Pick 'n Pay	0.28	0.32	0.30	0.29	0.06	0.17	1.00

Besides the correlations, a brief overview of the volatility over the ten year period is also considered. A plot of the log returns of the JSE Top40 Index is displayed in Figure 5.1 can be used to get an idea of the volatility in the South African market. The figure indicates that over the ten years of data there are periods of high volatility and periods of low volatility. For example, from about May 2008 to May 2009 there is a period of high volatility whereas the year 2005 was mostly a period of low volatility. In addition the effects of important events can be seen on the returns, such as the burst of the dot-com bubble in March 2000 and the 9/11 attacks on America in 2001.

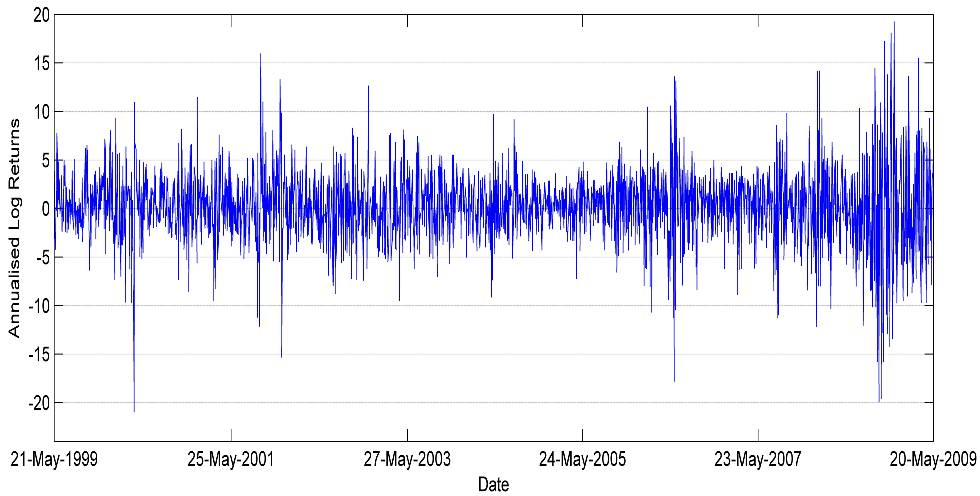


Figure 5.1: Plot of the Annualised Returns of the Top40 Index

5.1.1 Summary Statistics of the Share Data

Now that a broad overview of the share data has been discussed a more detailed discussion of the data follows. The summary statistics of the annualised log returns are shown in Table 5.3.

The sample means of all the returns of the shares are positive, however the medians are zero. This indicates that the data are asymmetrically distributed, which is verified by the non zero skewness. In fact all the distributions, except for that of ABSA, are positively skewed. A positively skewed distribution is preferable for investors as extreme returns have a greater probability of being positive rather than negative.

With regards to extreme returns, the difference between the minimum and maximum of each of the share returns is large and therefore a wide range of returns are observed. This is supported by the kurtosis, which is greater than 3 for each of the shares. Thus the distributions are more peaked than the normal distribution and have fatter tails so there is a greater probability of observing extreme returns.

Table 5.3: Shares - Summary Statistics

Statistic	**ABSA**	**FirstRand**	**Std Bank**	**Nedbank**
Mean	0.180	0.129	0.212	0.034
Median	0.000	0.000	0.000	0.000
Standard Deviation	5.433	5.449	5.428	5.414
Coefficient of Variation	30.246	42.401	25.623	158.361
Kurtosis	6.836	4.863	5.179	5.215
Skewness	-0.059	0.062	0.202	0.122
Minimum	-44.896	-31.437	-26.181	-27.138
Lower Quartile	-2.732	-3.030	-2.793	-3.100
Upper Quartile	2.944	3.205	3.106	2.915
Maximum	28.053	26.904	29.209	29.598
Jarque-Bera Statistic	1516.447	358.352	505.532	511.206

Statistic	**GoldFields**	**MurrayRob**	**PicknPay**
Mean	0.218	0.310	0.183
Median	0.000	0.000	0.000
Standard Deviation	8.287	6.160	4.942
Coefficient of Variation	37.982	19.867	26.968
Kurtosis	6.958	7.136	6.471
Skewness	0.423	0.012	0.401
Minimum	-39.528	-39.766	-24.142
Lower Quartile	-4.314	-2.793	-2.336
Upper Quartile	4.572	3.508	2.680
Maximum	62.259	41.270	34.731
Jarque-Bera Statistic	1687.258	1761.695	1306.980

In addition to the skewness and the kurtosis indicating non-normality, the Jarque-Bera statistics also indicate non normality. Each of the Jarque-Bera Statistics are much larger than 5.968 which is the 5% cut-off for testing the null hypothesis that the data are normally distributed. Hence the kurtosis, Jarque-Bera statistic and skewness all indicate that the share returns are not normally distributed.

Lastly the standard deviation and the coefficient of variation are considered. Gold Fields is the most risky as far as standard deviation is concerned but with regard to the coefficient of variation Nedbank is the most risky. This is most likely because the coefficient of variation adjusts the variation relative to the mean and Nedbank has the smallest mean.

5.2 General Description of the Exchange Rate Data

The second data set consists of daily exchange rates of the South African Rand (R) against five major foreign currencies from the 21st May 1999 to 20th May 2009, the names of which are given in Table 5.4. Each exchange rate is in the format of the number of Rands that are equivalent to one unit of foreign currency. These data were obtained from the South African Reserve Bank website
(http://www.resbank.co.za/economics/histdownload/histdownload.htm).

Table 5.4: Data Set 2 - Exchange Rates

Exchange Rates
Rand/Pound
Rand/Euro
Rand/United States of America $
Rand/Australian $
Rand/Yen

However the exchange rates may not be comparable over time due to outside interventions, usually due to government interference. Therefore a brief history of the South African government policy and their intervention in South African exchange rates is discussed over the ten year period of interest. Prior to 2000 there was substantial direct intervention in markets to ensure that exchange rates were in line with policy objectives. However in February 2000 an inflation targeting monetary policy framework was initiated with an inflation target of 3% to 6%, to be met within two years. Hence, from this point onwards, the South African Reserve Bank no longer intervened with the objective of influencing exchange rates and exchange controls were also relaxed. Of course government interventions in the other 5 regions (Euro zone, America, Japan, United Kingdom and Australia) should also be considered but these are not discussed here.

In addition to the impact government policy has on exchange rates, the governor of the South African Reserve Bank also plays a major role in decisions which affect South African exchange rates. Therefore exchange rates are likely to be more comparable over the period if there was a governor in place for most of the ten year period. Tito Mboweni had been governor from August 1999 to November 2009. Hence for most of the ten year period considered in the present study there was one governor. Therefore, from the perspective of South Africa and ignoring the other 5 regions, the exchange rates should be comparable over the ten years from May 1999 to May 2009 as government intervention in exchange rates was minimal because for most of the period the same policy and governor were in place.

Besides outside intervention, the correlations of the exchange rate returns are also considered and are shown in Figure 5.5. These correlations are noticeably higher than the correlations between the share returns.

Table 5.5: Correlation of the Exchange Rate Returns

	R/£	**R/€**	**R/US$**	**R/Aus$**	**R/Yen**
R/£	1.00	0.89	0.85	0.79	0.79
R/€	0.89	1.00	0.82	0.78	0.80
R/US$	0.85	0.82	1.00	0.71	0.87
R/Aus$	0.79	0.78	0.71	1.00	0.63
R/Yen	0.79	0.80	0.87	0.63	1.00

5.2.1 Summary Statistics of the Exchange Rate Data

Now that a broad overview of the exchange rate data has been provided a more detailed discussion of the data follows. The summary statistics of the annualised log returns of the exchange rates are shown in Table 5.6.

Table 5.6: Exchange Rates - Summary Statistics

Statistic	**R/£**	**R/€**	**R/US$**	**R/Aus$**	**R/Yen**
Mean	0.027	0.056	0.030	0.046	0.056
Median	-0.077	-0.099	-0.136	0.000	-0.092
Standard Deviation	2.778	2.773	2.958	2.654	3.477
Coefficient of Variation	103.973	49.746	97.343	58.053	62.196
Kurtosis	8.082	8.084	8.696	7.001	8.892
Skewness	0.600	0.618	0.675	0.310	0.527
Minimum	-14.328	-14.726	-18.506	-15.411	-21.726
Lower Quartile	-1.531	-1.526	-1.553	-1.452	-1.815
Upper Quartile	1.426	1.417	1.449	1.475	1.665
Maximum	23.373	23.961	26.379	18.026	28.524
Jarque-Bera Statistic	2821.997	2831.558	3544.657	1697.669	3711.841

The mean returns of the exchange rates are considerably less than the mean returns of the shares. This is because an exchange rate in itself is not an asset and therefore the share returns and exchange rate returns are not directly comparable. An investor would not purchase a currency and simply hold it. Typically an investor would purchase a bond in the desired currency or invest the money in the bank in the desired currency and so forth. Hence the returns would usually be greater due to the coupons received, the interest earned and income received.

Assuming purchasing power parity holds, it is expected that on average all five exchange rates should increase over the ten year period. This is because on average the inflation rate over the ten year period in South Africa is greater than the inflation rates in the United States of America, Euro zone, United Kingdom, Japan and Australia. Although the means suggest an increase in exchange rates over the period, the medians of four of the five exchange rates are negative.

All the exchange rate return distributions appear to be non normal as they are positively skewed, have a kurtosis greater than three and the Jarque-Bera statistics indicate a non normal distribution. The standard deviations of the exchange rate returns are about half of what the standard deviations of the share returns are, but the coefficients of variation are typically greater due to the small means. The ranges are also much smaller. Therefore it appears that the exchange rate returns are less risky than the share returns. However, as previously mentioned, the two types of financial data are not really directly comparable.

Chapter 6

Results - ARMA(p,q) and VAR(p) Models

It is important to get an idea of the conditional mean of the log returns $\boldsymbol{z_t}$ before applying the procedure in Section 4.2.1 to choose the number of AR parameters p and the number of MA parameters q in the appropriate ARMA(p,q) and VAR(p) models. This is necessary as a process should not be mindlessly applied without understanding the relevant features of the underlying data. To gain an understanding of the conditional mean of the returns $\boldsymbol{z_t}$, the sample autocorrelations, partial autocorrelations and possibly any cross autocorrelations and partial cross-correlations should be examined. However the cross autocorrelations and partials are not discussed due to limited space.

Once the bigger picture has been considered the best values for p and q can be selected for the ARMA(p,q) and VAR(p) models. These models can then be fitted to the data to determine the parameter estimates. Each of these fitted models needs to be tested to ensure that the model fits the data adequately. If the model fit is not suitable then the model needs to be adjusted accordingly.

As the conditional mean is not the focus of this thesis, only a quick overview of the steps involved in selecting values for p and q are given in this chapter.

6.1 Model Representation

Before discussing any model results it is necessary to define some notation and the model representations.

Recall from Figure 2.1 that $\boldsymbol{e_t} = (e_{1t}, e_{2t}, \cdots, e_{Nt})$ is an $N \times 1$ vector of the residuals of the ARMA(p,q) models or the VAR(p) model at time t. In the case of the ARMA(p,q) models e_{it} is the residual at time t of the ARMA(p,q) model fitted to the i^{th} series of returns, z_{it} for $i = 1, 2, \cdots N$ and $t = 1, 2, \cdots T$. Thus an ARMA(p,q) model is fitted to each series separately so that there are N ARMA(p,q) models which may have different values of p and q. Therefore an ARMA(p,q) model fitted to the i^{th} series of returns $\boldsymbol{z_t}$ can be represented as

$$z_{it} = c_i + \rho_{i1}\, z_{i,t-1} + \rho_{i2}\, z_{i,t-2} + \cdots + \rho_{ip}\, z_{i,t-p} + \tag{6.1}$$
$$e_{it} + \theta_{i1}\, e_{i,t-1} + \theta_{i2}\, e_{i,t-2} + \cdots + \theta_{iq}\, e_{i,t-q} \tag{6.2}$$

where c_i is a constant, $\rho_{i1}, \rho_{i2}, \cdots \rho_{ip}$ are the AR parameters and $\theta_{i1}, \theta_{i2}, \cdots \theta_{iq}$ are the MA parameters.

In the context of the VAR(p) model, e_{it} is the i^{th} series residual at time t. Unlike the ARMA(p,q) models, the VAR(p) model is fitted to all N series of returns $\boldsymbol{z_t}$ simultaneously. Therefore a VAR(p) model can be represented as

$$\boldsymbol{z_t} = \boldsymbol{c} + \boldsymbol{R_1}\, \boldsymbol{z_{t-1}} + \boldsymbol{R_2}\, \boldsymbol{z_{t-2}} + \cdots + \boldsymbol{R_p}\, \boldsymbol{z_{t-p}} + \boldsymbol{e_t}$$

where $\boldsymbol{c}$ is an $N \times 1$ vector of constants and $\boldsymbol{R_1}, \boldsymbol{R_2}, \cdots \boldsymbol{R_p}$ are $N \times N$ matrices containing the VAR parameters.

6.2 Autocorrelations and Partial Autocorrelations of the Returns $\boldsymbol{z_t}$

From the representation of the ARMA(p,q) model it can be observed that the number of significant autocorrelations of a series is approximately the value of q and the number of significant partial autocorrelations of a series is approximately the value of p. However this is only approximate because a stationary AR process can be represented as an infinite MA process and visa versa. Therefore using the autocorrelations and partial autocorrelations to determine p and q is not a strict rule but simply a guideline.

On the other hand, a VAR(p) model is somewhat different to an ARMA(p,q) model because to determine p the partial autocorrelations between series need to be considered in addition to the partial autocorrelations of each series. Note that it is of little use to examine the autocorrelations in the context of a VAR(p) model.

Consequently plots of the autocorrelations and partial autocorrelations of each series and the partial autocorrelations between each series should be examined to help determine p and q for the ARMA(p,q) and VAR(p) models. In spite of this, only figures of the autocorrelations of each series are included in this study. All the plots are examined but only a few key figures are displayed. The first reason for this is that plots of the sample partial autocorrelations are almost identical to the plots of the sample autocorrelations. Secondly, plots of the cross partial autocorrelations are excluded to keep the total number of plots displayed to a minimum. Therefore the autocorrelation plots are simply there to give some insight into how the conditional mean can be modelled.

Plots of the sample autocorrelations of the share and exchange rate returns at various lags are found in Figures 6.1 and 6.2 respectively. The two horizontal lines above and below the x-axis give an approximate 95% confidence interval under the null hypothesis that the autocorrelations are zero. Thus any sample autocorrelations which do not lie between these lines are considered to be significant and thus need to be modelled.

Although some of the share and exchange rate returns have one or two significant sample autocorrelations at fairly high lags these are not a concern. This is because significant sample autocorrelations at fairly high lags are probably spurious and due to a few large positive or negative returns which are that particular number of lags apart.

With regard to the share data in Figure 6.1 it can be observed that the sample autocorrelation of most share returns are significant at the first two lags and a number are also significant at lags 3 and 4. What is interesting to note is that most of the sample autocorrelations at lag 1 are positive and then most are negative from lags 2 to 4. Thus share returns appear to display some momentum in the very short term, that is today's return is positively related to the previous days return.

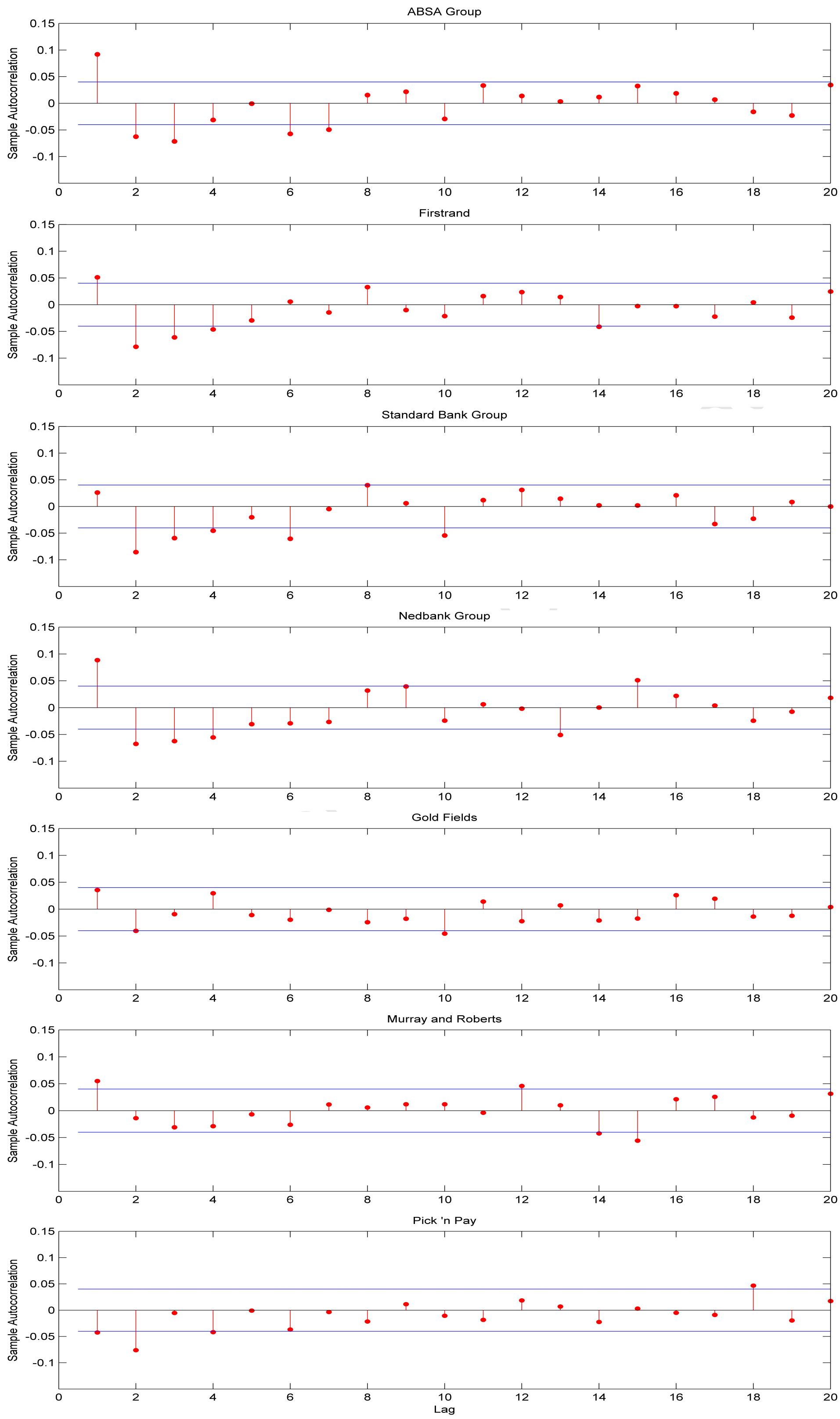


Figure 6.1: Sample Autocorrelation of the Share Returns at Various Lags

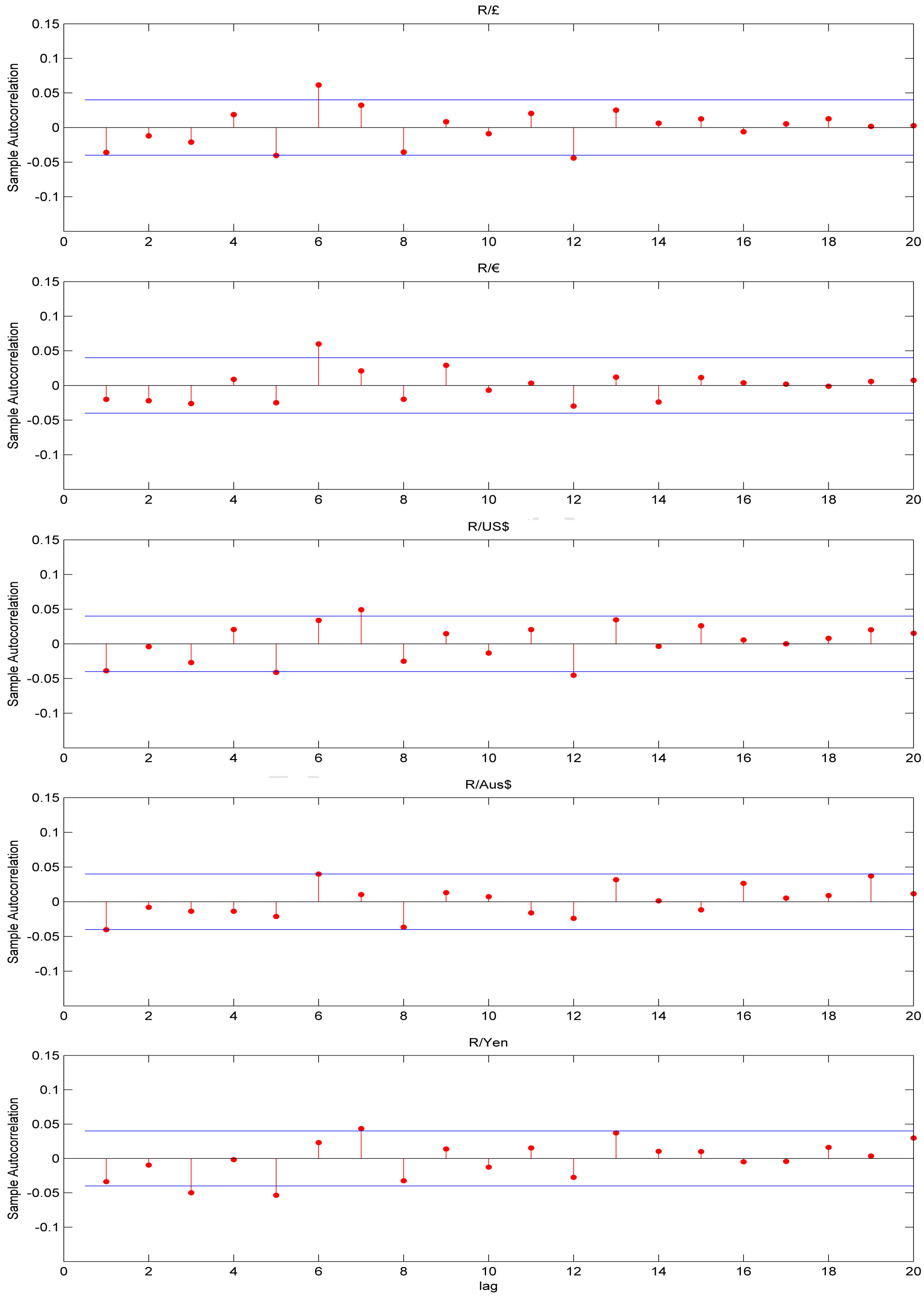


Figure 6.2: Sample Autocorrelation of the Exchange Rate Returns at Various Lags

In contrast to the share returns, the exchange rate returns have very few significant sample autocorrelations at lags 1 to 4. Another difference is that the sample autocorrelations of the share returns at lag 1 are mostly positive but they are all negative for the exchange rate returns. Additionally at lag 1 all of the autocorrelations of the exchange rate returns are very close to being significant but none actually are. However at lags 5 and 6 most of the exchange rate returns are either significant or very close to being significant. The sample autocorrelations at lag 5 are all negative but at lag 6 they are all positive. This is interesting because significant autocorrelation at lag 5 in the context of daily returns means that the exchange rate return today is related to the return a week ago, that is the return last Monday can help to predict the return on the coming Monday and so forth.

Furthermore, it is worth noting that despite the fact that markets are generally hypothesised to be arbitrage-free significant sample autocorrelations can be observed in both data sets. Possible reasons for this were discussed earlier.

6.3 Optimal Values of p and q

Now that a broad overview of the conditional means has been discussed, the values of p and q in the ARMA(p,q) and VAR(p) models for the share returns and the exchange rate returns can be determined using the procedure described in Section 4.2.1. These values are constrained to be a maximum of 6 each. The reason a maximum value of 6 is chosen for q is that in Figures 6.1 and 6.2 there are very few significant autocorrelations at lags greater than 6 and in any case the ones which are significant at lags greater than 6 are considered to be spurious. Since the sample autocorrelations and partial autocorrelations are very similar, the maximum value allowed when choosing p is also 6.

The optimal values of p and q, determined using the procedure in Section 4.2.1 but up to a maximum of 6 for each series of share and exchange rate returns for the ARMA(p,q) models are given in Tables 6.1 and 6.2 respectively.

In spite of the fact that the share returns generally had significant partial autocorrelations at lower lags and the exchange rate returns generally had significant sample partial autocorrelations at lags 5 and 6, the values of p for most shares are fairly close to six but for most exchange rates they are smaller. However, most of the values of q for both the exchange rate and share returns are close to 6. The discrepancies between the values of p and q compared to the partials and autocorrelations are because of the relationship between the AR and the MA terms in that a stationary AR model can be represented as an infinite MA model and visa versa.

Table 6.1: Shares - Best fitting ARMA(p,q) Model

	p	q
ABSA	4	6
FirstRand	6	5
Std Bank	5	5
Nedbank	4	3
GoldFields	6	6
MurrayRob	4	5
Pick 'n Pay	3	3

Table 6.2: Exchange Rates - Best fitting ARMA(p,q) Model

	p	q
R/£	2	6
R/€	0	6
R/US$	2	4
R/Aus$	3	3
R/Yen	5	5

For the VAR(p) models, the optimal value of p, again determined using the procedure in Section 4.2.1 but up to a maximum p value of 6, is found to be 6 for both the share and exchange rate data sets.

6.4 Parameter Estimates

Since the best values for p and q (up to a maximum of 6) have been determined, the ARMA(p,q) and VAR(p) models can be fitted to the returns using the number of parameters chosen. The parameters are estimated in MATLAB by maximising the log likelihood.

The resulting parameter estimates for each of the ARMA(p,q) models are found in Appendix B in Sections B.1 and B.2 . In addition the estimates of the model parameters of the two VAR(p) models can be found in Appendix B in Sections B.3 and B.4.

The fit of each of the models will now be discussed. However the actual values of the parameter estimates are not discussed as the conditional mean models are not the focus of the thesis and there would, in any case, be a large number of parameters to discuss.

6.5 Testing the Fit of the Models

Once the conditional mean models have been fitted the residuals need to be tested to determine whether the conditional mean has been adequately modelled. Two tests are performed to determine if the models fit adequately and these are now discussed.

6.5.1 Ljung-Box Q-Statistic

If the conditional mean has been adequately modelled then the residuals should have no significant autocorrelations. The autocorrelations of the residuals are tested using a Ljung-Box Q-Statistic up to lag 10. The null hypothesis of this test is that all the autocorrelations up to a lag of 2 weeks, that is 10 working days, are zero and the alternative hypothesis is that at least one of the autocorrelations in the first ten lags is not zero.

The reason for choosing 10 lags is that there is a compromise between the power of the test and the number of lags tested. Choosing a low number of lags will ensure that the test has a high power but then the autocorrelations at higher lags are not tested. However choosing a high number of lags results in a test with low power but the autocorrelations at higher lags will also be tested. The reason that the results can be misleading if a high number of lags is chosen is that the Q-Statistic sums up the squared sample autocorrelations up to the chosen number of lags, in this case 10. Therefore if the chosen number of lags is large and most autocorrelations are small except for one or two then this will result in a small Q-Statistic which is likely to be insignificant when it should in fact be significant.

The Q-Statistics of the residuals from the fitting the ARMA(p,q) models to the share and exchange rate returns are given in Tables 6.3 and 6.4 respectively along with the associated probabilities of observing a larger Q-Statistic. These probabilities are referred to as p-values.

The VAR(p) model is somewhat different in that it is a multivariate model and therefore each series of residuals can no longer be associated with one share or exchange rate. Therefore each series of residuals will be represented by a number, that is series 1 to series 6. The Q-Statistic of the share and exchange rate residuals obtained by fitting a VAR(6) model and the associated probabilities of observing a larger Q-Statistic are given in Tables 6.5 and 6.6. Likewise, the probability is called a p-value.

Table 6.3: Shares - Q-Statistics and p-values for the optimal ARMA(p,q) models

	Q_{10}	p-value
ABSA	3.3497	0.9720
FirstRand	2.4119	0.9921
Std Bank	5.3626	0.8657
Nedbank	6.3293	0.7869
GoldFields	2.8067	0.9856
MurrayRob	1.3230	0.9994
PicknPay	2.5057	0.9908

Table 6.4: Exchange Rates - Q-Statistics and p-values for the optimal ARMA(p,q) models

	Q_{10}	p-value
R/£	2.7975	0.9858
R/€	3.5110	0.9667
R/US$	13.1305	0.2165
R/Aus$	6.1345	0.8038
R/Yen	9.0388	0.5284

In all four Tables, namely 6.3, 6.4, 6.5 and 6.6, the tail probabilities are all greater than 0.2. Hence even at the 20% significance level the null hypothesis that all the autocorrelations up to lag 10 are zero cannot be rejected.

Table 6.5: Shares - Q-Statistics and p-values for the VAR(6) model

	Q_{10}	p-value
Series 1	7.2929	0.6975
Series 2	2.5828	0.9896
Series 3	11.2320	0.3397
Series 4	7.5129	0.6763
Series 5	8.4009	0.5897
Series 6	1.1948	0.9996
Series 7	5.3074	0.8697

Table 6.6: Exchange Rates - Q-Statistics and p-values for the VAR(6) model

	Q_{10}	p-value
Series 1	4.7982	0.9042
Series 2	4.6969	0.9105
Series 3	7.3336	0.6936
Series 4	4.2474	0.9355
Series 5	6.7298	0.7507

Therefore based on this test, it appears that each of the ARMA(p,q) and VAR(p) models adequately models the conditional mean.

6.5.2 Significance of the Model Parameters

For both the ARMA(p,q) and the VAR(p) models a t-statistic, calculated under the null hypothesis that the particular parameter is zero, can be found for each model parameter. If $t_{observed}$ is the value of the t-statistic then the p-value is the probability of observing a more extreme value than $t_{observed}$ given that the true parameter is zero. Therefore if a t-statistic is significant then that model parameter should be included in the model as the null hypothesis that the parameter is zero is rejected.

The p-values associated with each parameter for the ARMA(p,q) models are given in the Appendix in Sections B.1 and B.2 and for the VAR(p) models in Appendix Sections B.3 and B.4. Besides the constants c_i's , see equation (6.2), most of the t-statistics of the ARMA(p,q) parameters are significant at the 5% significance level. However, with regard to the two VAR(p) models, many of the t-statistics are not significant at the 5% or even at the 10% significance level. This suggests that fitting a VAR model to the returns is probably unnecessary. However, even though many of the parameters are not significant, the value of p used when fitting the VAR(p) models will still be 6. This is because the aim of this step in the calculations is simply to remove the conditional mean and not to find the best model. In any case the Q-statistic suggests that this has been achieved.

Since the values of p and q selected are suitable, from this point onwards when referring to the VAR(p) model, p is assumed to be 6 and for the ARMA(p,q) models the values of p and q are assumed to be those in Tables 6.3 and 6.4, unless stated otherwise.

Chapter 7

Results - Parameters of the O-GARCH, O-EWMA and O-SV Models

The previous chapter dealt with finding the optimal values of p and q for the ARMA(p,q) and VAR(p) models fitted to the the log returns $\boldsymbol{z_t}$ for $t = 1, 2, \cdots T$. The resulting estimate of the conditional mean using the optimal values of p and where appropriate q is $\boldsymbol{\hat{\mu}_t}$. Therefore the conditional mean residuals can then be calculated as $\boldsymbol{\hat{e}_t} = \boldsymbol{z_t} - \boldsymbol{\hat{\mu}_t}$. These steps are outlined in Figure 2.1 along with a number of steps which follow. This chapter deals with the remaining steps in Figure 2.1 which are now briefly recapped.

After calculating the sample residuals $\boldsymbol{\hat{e}_t}$, the next step in the model is to adjust the sample residuals using the matrix $\boldsymbol{B}$ so that $\boldsymbol{y_t} = \boldsymbol{B}\boldsymbol{\hat{e}_t}$. Recall that the two choices for $\boldsymbol{B}$ are $\boldsymbol{B} = \boldsymbol{I}$ and $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$. The standardised residuals $\boldsymbol{y_t}$ are used to construct the principal component scores $\boldsymbol{x_t}$ via an $N \times N$ matrix of factor weights $\boldsymbol{A}$ such that $\boldsymbol{x_t} = \boldsymbol{A^T}\boldsymbol{y_t}$. The columns of $\boldsymbol{A}$ are the eigenvectors of $\boldsymbol{\hat{V}}$, the covariance matrix of $\boldsymbol{y_t}$.

Before the principal component scores are standardised and a conditional volatility model is fitted, the number of series of principal components to include in the model needs to be determined. This chapter discusses an appropriate choice of h, the number of series to include, for the share and exchange rate data sets. This is selected by examining the proportion of variance explained by each series of principal component scores. Once h has been selected the factors are standardised using the matrix $\boldsymbol{W^{(h)}}$ to give $\boldsymbol{u_t^{(h)}} = \boldsymbol{W^{(h)}}\boldsymbol{x_t^{(h)}}$. The matrix $\boldsymbol{W^{(h)}}$ may either be $\boldsymbol{W^{(h)}} = \boldsymbol{I}$ or

$\boldsymbol{W}^{(h)} = (\boldsymbol{\Lambda}^{(h)})^{-\frac{1}{2}}$. Finally, one of three possible conditional volatility models are fitted to the standardised principal component scores $\boldsymbol{u_t^{(h)}}$.

To summarise, this chapter discusses the factor weights $\boldsymbol{A}$, the proportion of variance explained by each series of principal component scores and the model parameters of the O-GARCH, O-EWMA and O-SV models.

7.1 Factor Weights and Proportion of Variance Explained

It is assumed that the conditional mean residuals $\boldsymbol{\hat{e}_t}$ have been calculated and adjusted, that is $\boldsymbol{y_t} = \boldsymbol{B}\boldsymbol{\hat{e}_t}$, such that either $\boldsymbol{B} = \boldsymbol{I}$ or $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$. Therefore this section is concerned with the matrix of factor weights $\boldsymbol{A}$ and the number of series of principal component scores to include in the model.

Recall that the factor weights $\boldsymbol{A}$ are constructed so that the columns of $\boldsymbol{A}$ consist of the eigenvectors of $\boldsymbol{\hat{V}}$ which is the unconditional sample covariance matrix of $\boldsymbol{y_t}$. Additionally $\boldsymbol{A}$ is constructed such that the first column of $\boldsymbol{A}$ is the eigenvector corresponding to the largest eigenvalue and the second column corresponds to the second largest eigenvalue and so forth. Hence the eigenvalue of $\boldsymbol{\hat{V}}$ which corresponds to the i^{th} column of $\boldsymbol{A}$ is the sample variance of the i^{th} series of principal component scores.

Thus the proportion of the total variance explained by the i^{th} series of the principal component scores is the i^{th} eigenvalue divided by the sum of the eigenvalues. Therefore the cumulative portion of variance explained by the first h series of principal component scores can be calculated by summing the first h proportions. These cumulative proportions can help to determine h, the number of series of principal component scores included in the model (Alexander, 2000). The cumulative proportions are given for the share and exchange rate data sets in Tables 7.1 and 7.2 respectively. The actual eigenvalues used to calculate these cumulative proportions can be found in the Appendix in Section C.1. In addition the factor weights $\boldsymbol{A}$ can also be found in the Appendix in Section C.2.

Once a value of h has been selected using this method, each of the complete orthogonal factor models need to be fitted using this value of h to determine whether the resulting conditional covariance estimates for $\boldsymbol{z_t}$ are reasonable. The models need to be fitted for each of the conditional mean models, $\boldsymbol{B}$

Table 7.1: Shares - Cumulative Proportion of the Variance Explained by the Principal Component Scores

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$	$\boldsymbol{B}$	Number of Principal Components Included						
		1	2	3	4	5	6	7
ARMA	$\boldsymbol{I}$	0.36	0.64	0.77	0.85	0.91	0.96	1.00
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.43	0.58	0.70	0.82	0.89	0.95	1.00
VAR	$\boldsymbol{I}$	0.36	0.64	0.77	0.85	0.91	0.96	1.00
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.44	0.58	0.71	0.82	0.89	0.95	1.00

Table 7.2: Exchange Rates - Cumulative Proportion of the Variance Explained by the Principal Component Scores

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$	$\boldsymbol{B}$	Number of Principal Components Included				
		1	2	3	4	5
ARMA	$\boldsymbol{I}$	0.83	0.92	0.95	0.98	1.00
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.83	0.92	0.95	0.98	1.00
VAR	$\boldsymbol{I}$	0.84	0.92	0.95	0.98	1.00
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.84	0.92	0.95	0.98	1.00

and $\boldsymbol{W}$.

Hence, in the case of the share data set, the cumulative proportion of variance explained by the principal component scores suggest that choosing $h = 5$, that is including 5 principal components in the model, is probably appropriate. This is because about 90% of the variance of the share returns are explained by the first 5 of the 7 principal components.

On the other hand, in the case of the exchange rate data set, the cumulative proportion of the variances explained by the principal component scores suggest that choosing $h = 2$, that is using 2 principal components in the models, may be sufficient. However choosing $h = 2$ may be problematic in terms of accurately modelling the conditional covariances as only 2 independent sources of risk are included in such a model.

In spite of the fact that the results summarised in Tables 7.1 and 7.2 suggest that the majority of the variance can be explained by the first few series of principal component scores a preliminary study suggests that choosing $h = 5$ and $h = 2$ are not suitable. To illustrate this a plot of the conditional correlations of a pair of share returns and a pair of exchange rate returns are considered.

Thus the models fitted to the share returns are tested using the first 5 series of principal component scores instead of all 7. The resulting O-GARCH conditional correlation estimates of the returns of ABSA and FirstRand are given in Figure 7.1 where the conditional mean model is ARMA, $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ and $\boldsymbol{W} = \boldsymbol{I}$. The choice of the conditional mean model, $\boldsymbol{B}$ and $\boldsymbol{W}$ in the plot appear to have little impact on the resulting plot. It is clear from Figure 7.1 that these estimates are not appropriate as all of the conditional correlation estimates are much greater than the unconditional sample correlation of 0.59 (see Table 5.2). The corresponding plots for the O-EWMA and O-SV models have not been included as they are similar in the sense that the level of the conditional correlations are very similar and therefore the resulting estimates are also unreasonable. Likewise, the conditional correlation estimates of the other pairs of share returns did not appear reasonable.

A plot of each of the O-GARCH, O-EWMA and O-SV model estimates of the conditional correlation estimates of the returns of ABSA and FirstRand can be found in the next chapter in Figure 8.2 when $h = 7$. Similarly in this figure the conditional mean model is an ARMA model, $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ and

$\boldsymbol{W} = \boldsymbol{I}$. The estimates using $h = 7$ are more reasonable and are discussed in more detail in the that chapter.

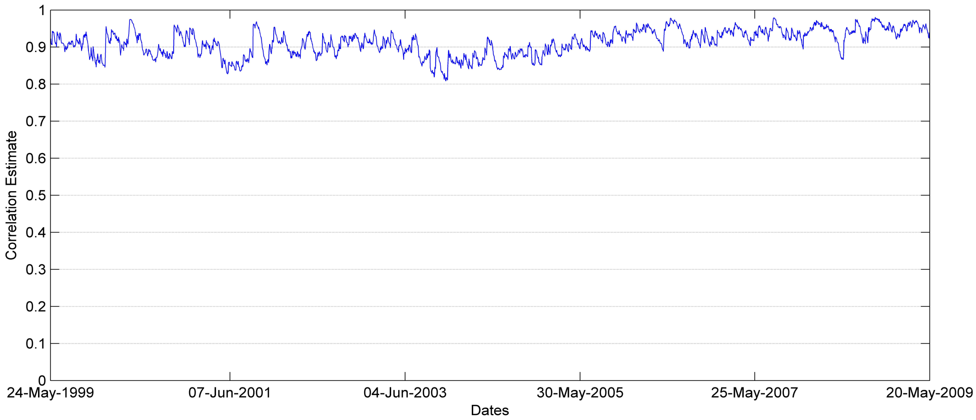


Figure 7.1: Plot of the Conditional Correlation between the Returns of ABSA and FirstRand when $h = 5$

The models applied to the exchange rate data set are tested using $h = 2$, that is 2 principal components are included in the models. Despite this the plots of the resulting conditional correlation estimates of the R/£ and R/€ returns are not included as it appears that the conditional correlation is 1 throughout the time period. This is also the case when $h = 3$. Although the estimates do not visibly vary on the plots, they do in fact vary slightly but are always extremely close to or equal to 1.

Thus instead of plotting the conditional correlations of the models where 2 or 3 principal components are included, the estimates are plotted where the model includes 4 principal components as these estimates can visibly be seen to vary on a plot. The O-GARCH conditional correlation estimates of the returns of the R/£ and R/€ are given in Figures 7.2 where $\boldsymbol{B} = \boldsymbol{I}$ and 7.3 where $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$. In both these plots the conditional mean model used is ARMA and $\boldsymbol{W} = \boldsymbol{I}$. Similar to the share data sets, the choice of the conditional mean model and $\boldsymbol{W}$ make little difference to the results. On the other hand it is clear from the two plots that the choice of $\boldsymbol{B}$ makes a noticeable difference to the conditional correlation estimates in the exchange rate data set.

It is clear from Figure 7.2 that the conditional correlation estimates are not reasonable given that the unconditional sample estimate is 0.89 (see Table

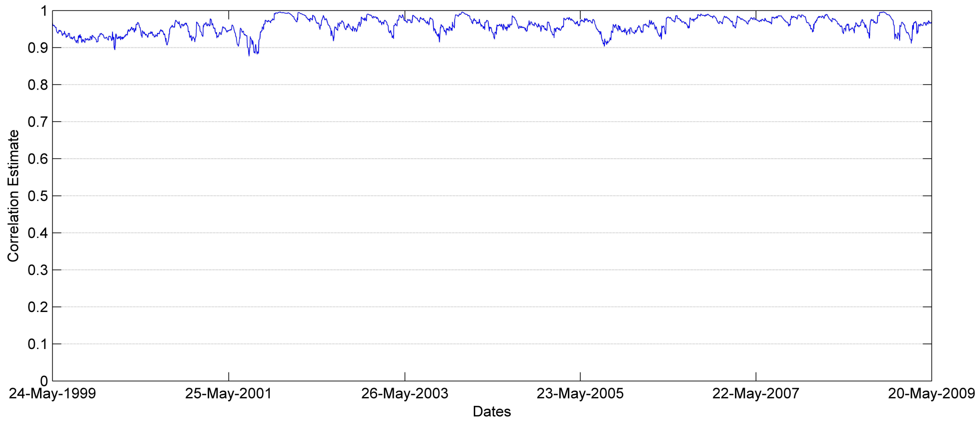


Figure 7.2: Plot of the Conditional Correlation between the Returns of the R/£ and R/€ when $\boldsymbol{B} = \boldsymbol{I}$ and $h = 4$

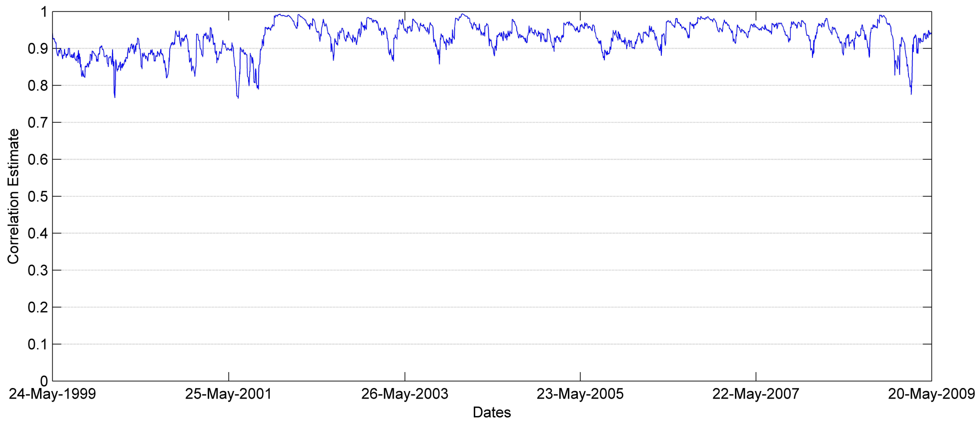


Figure 7.3: Plot of the Conditional Correlation between the Returns of the R/£ and R/€ when $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ and $h = 4$

5.5). However, it is interesting to note that in Figure 7.3 although the estimates still do not seem suitable, they are more reasonable. This indicates that using $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ may be preferable to using $\boldsymbol{B} = \boldsymbol{I}$ as it appears to result in more reasonable estimates for the conditional correlations.

The plots for the O-EWMA and O-SV models are not included as they are similar in the sense that the level of the conditional correlations are very similar. Likewise, the conditional correlation estimates of the other pairs of exchange rate returns did not appear reasonable.

A plot of each of the O-GARCH, O-EWMA and O-SV model estiamtes of the conditional correlation estimates of the returns of the R/£ and R/€ Exchange Rates can be found in the next chapter in Figure 8.4 when $h = 5$ and the conditional mean model is ARMA, $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ and $\boldsymbol{W} = \boldsymbol{I}$. The estimates using $h = 5$ are more reasonable and are discussed in more detail in the next chapter.

Thus for the two data sets tested, it appears that the suggestion made by Alexander (2000) that it is preferable to use only the first few series of principal component scores is not applicable. Since the conditional correlation estimates are not reasonable when using some but not all of the series of principal component scores, the model results from this point onwards are for the models which include all the series. In other words, for the share data set $h = 7$ and for the exchange rate data set $h = 5$ from this point onwards.

7.2 Parameter Estimates

The next step is to standardise the principal component scores to give $\boldsymbol{u_t} = \boldsymbol{W}\boldsymbol{x_t}$ where the two choices for $\boldsymbol{W}$ are $\boldsymbol{W} = \boldsymbol{I}$ and $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$. Since $h = N$, in other words all possible series are included in the model, $\boldsymbol{W}^{(\boldsymbol{h})}$ is equivalent to $\boldsymbol{W}$, $\boldsymbol{u_t^{(h)}}$ is equivalent to $\boldsymbol{u_t}$ and $\boldsymbol{x_t^{(h)}}$ is equivalent to $\boldsymbol{x_t}$. Therefore it is no longer necessary to include the superscript $^{(h)}$. Thus the final step is to fit one of the three conditional volatility models to the standardised principal component scores. The resulting parameter estimates of each of these models are now discussed.

7.2.1 O-GARCH Parameter Estimates

The first conditional volatility model fitted to the standardised principal component scores is the GARCH(1,1) model. To recap the GARCH(1,1) model equation is

$$\sigma_{it}^2 = Var(u_{it}|F_{t-1}) \quad = \quad \alpha_{i,0} + \alpha_{i,1}\, u_{i,t-1}^2 + \beta_i\, \sigma_{i,t-1}^2$$

where $\alpha_{i,1}$ measures the extent of the markets reaction and β_i represents how persistent the volatitlity is. Estimates of the O-GARCH parameters α_{i0}, α_{i1} and β_i for the share data sets can be found in Tables 7.3, 7.4 and 7.5 respectively and for the exchange rate data set in Tables 7.6, 7.7 and 7.8 respectively.

Furthermore each estimate has an associated t-statistic which is calculated under the null hypothesis that the true parameter is zero. Hence a p-value can be calculated for each t-statistic which is the probability of observing a more extreme t-statistic. These p-values are all smaller than 0.02, except for the p-value associated with α_{30} in the exchange rate data set where the conditional mean model is a VAR model, $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ and $\boldsymbol{W} = \boldsymbol{I}$ which has a large p-value associated with it. Hence the null hypothesis that the parameters are zero can be rejected for all the parameters except for the one case mentioned. This is interesting as many of the estimates of α_{i0} are fairly close to zero, especially in the case of the exchange rate data set. This implies that the standard error of these parameter estimates are very small.

Table 7.3: Shares - Estimates of the GARCH Parameters α_{i0}

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	Number of the Principal Component						
			1	2	3	4	5	6	7
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	3.6376e-5	1.5279e-5	3.2346e-5	1.5399e-5	1.1535e-6	1.6163e-6	2.1335e-6
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.0258	0.0141	0.0637	0.0492	0.0057	0.0084	0.0133
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.0827	0.0161	0.0411	0.0326	0.0028	0.0035	0.0048
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.0272	0.0157	0.0478	0.0407	0.0059	0.0082	0.0133
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	3.6051e-5	1.4390e-5	4.0447e-5	1.6551e-5	1.0045e-6	1.2282e-6	1.5731e-6
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.0258	0.0135	0.0813	0.0532	0.0054	0.0069	0.0105
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.0864	0.0140	0.0048	0.0411	0.0025	0.0029	0.0036
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.0282	0.0136	0.0056	0.0510	0.0053	0.0069	0.0104

Table 7.4: Shares - Estimates of the GARCH Parameters α_{i1}

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	Number of the Principal Component						
			1	2	3	4	5	6	7
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	0.1069	0.0736	0.1341	0.1344	0.0220	0.0347	0.0626
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.1069	0.0737	0.1343	0.1348	0.0223	0.0353	0.0632
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.1083	0.0715	0.1132	0.1154	0.0225	0.0357	0.0629
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.1083	0.0715	0.1132	0.1154	0.0225	0.0357	0.0629
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	0.1057	0.0729	0.1499	0.1456	0.0200	0.0327	0.0564
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.1058	0.0730	0.1500	0.1459	0.0203	0.0332	0.0575
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.1073	0.0659	0.0339	0.1321	0.0204	0.0333	0.0570
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.1073	0.0659	0.0339	0.1321	0.0204	0.0333	0.0570

Table 7.5: Shares - Estimates of the GARCH Parameters β_i

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	Number of the Principal Component						
			1	2	3	4	5	6	7
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	0.8688	0.9131	0.8095	0.8218	0.9730	0.9571	0.9247
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.8687	0.9130	0.8091	0.8211	0.9723	0.9563	0.9236
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.8662	0.9133	0.8441	0.8475	0.9719	0.9561	0.9240
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.8662	0.9133	0.8441	0.8475	0.9719	0.9561	0.9240
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	0.8695	0.9148	0.7766	0.8083	0.9755	0.9608	0.9340
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.8694	0.9146	0.7763	0.8077	0.9745	0.9598	0.9323
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.8655	0.9211	0.9612	0.8216	0.9745	0.9597	0.9328
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.8655	0.9211	0.9612	0.8216	0.9745	0.9597	0.9328

Table 7.6: Exchange Rates - Estimates of the GARCH Parameters α_{i0}

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	Number of the Principal Component				
			1	2	3	4	5
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	6.3204e-6	4.483e-7	2.000e-7	2.000e-7	2.000e-7
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	1.1132e-2	8.052e-3	7.228e-3	9.911e-3	5.056e-3
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	4.5706e-2	3.1706e-3	8.7842e-4	1.5238e-3	5.6135e-4
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	1.0982e-2	7.6612e-3	4.6255e-3	1.0884e-2	5.9550e-3
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	7.1187e-6	3.4998e-7	2.0000e-7	2.0000e-7	2.0000e-7
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	1.2717e-2	6.5669e-3	7.6682e-3	1.0592e-2	5.6547e-3
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	5.2831e-2	2.5163e-3	2.0000e-7	1.5642e-3	5.8906e-4
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	1.2649e-2	6.1099e-3	5.0933e-3	1.1550e-2	6.4189e-3

Table 7.7: Exchange Rates - Estimates of the GARCH Parameters α_{i1}

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	Number of the Principal Component				
			1	2	3	4	5
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	0.0982	0.0609	0.0421	0.0398	0.0562
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.0983	0.0617	0.0406	0.0398	0.0427
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.0974	0.0613	0.0342	0.0468	0.0424
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.0974	0.0613	0.0342	0.0468	0.0423
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	0.0931	0.0597	0.0462	0.0382	0.0564
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.0933	0.0598	0.0448	0.0383	0.0438
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.0925	0.0596	0.1046	0.0449	0.0413
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.0925	0.0595	0.0397	0.0449	0.0413

Table 7.8: Exchange Rates - Estimates of the GARCH Parameters β_i

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	Number of the Principal Component				
			1	2	3	4	5
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	0.8971	0.9299	0.9493	0.9506	0.9264
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.8969	0.9289	0.9522	0.9507	0.9527
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.8981	0.9295	0.9613	0.9426	0.9522
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.8981	0.9295	0.9613	0.9426	0.9522
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	0.8994	0.9338	0.9449	0.9518	0.9254
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.8991	0.9331	0.9476	0.9516	0.9509
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.9001	0.9337	0.8954	0.9439	0.9527
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.9001	0.9337	0.9554	0.9439	0.9527

Furthermore there are a number of observations which can be made with regards to the parameter estimates themselves. It is evident in both data sets that the persistence of the conditional volatility, β_i, is fairly high and the extent of the markets reaction, α_{i1}, to the observed returns are fairly low. Thus the conditional volatility at a point in time dominates the conditional volatility at the next time step.

Another observation relates to the choice of $\boldsymbol{W}$. Recall the suggestion made by Bongers (2008) that choosing $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ should reduce the square error in the parameter estimates of a GARCH model. However for a given conditional mean model and choice of matrix $\boldsymbol{B}$, choosing either $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ or $\boldsymbol{W} = \boldsymbol{I}$ give very similar parameter estimates for α_{i1} and β_i. The estimates appear to be even closer when $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ than when $\boldsymbol{B} = \boldsymbol{I}$, in fact in this case they are often identical up to the fourth decimal place. Therefore this seems to suggest that the recommendation made by Bongers (2008) does not make much of a difference.

7.2.2 O-EWMA Parameter Estimates

The second conditional volatility model fitted to the standardised principal component scores is the EWMA model or IGARCH(1,1) model without drift. Recall that two of the defining equations for this conditional volatility model are

$$\begin{aligned} u_{it} &= \sigma_{it}\,\epsilon_{it} \\ \sigma_{it}^2 = Var(u_{it}|F_{t-1}) &= \pi_i \sigma_{i,t-1}^2 + (1-\pi_i) u_{i,t-1}^2. \end{aligned}$$

The O-EWMA parameter π_i is comparable to the O-GARCH parameter β_i and $1-\pi_i$ in the O-EWMA model is comparable to the O-GARCH parameter α_{i1} where $0 < \pi_i < 1$. Thus one might expect the O-EWMA parameter π_i to have a similar value to the O-GARCH parameter β_i and $1-\pi_i$ to have a similar value to the O-GARCH parameter α_{i1}.

Estimating the parameter π_i is tricky as certain values of the parameter π_i in the set $0 < \pi_i < 1$ induce autocorrelation into the squared residuals ϵ_{it}^2 (see the system of equations (3.2)). Therefore the idea is to try and select a value of π_i which maximises the likelihood function over a range of values for π_i which do not induce autocorrelation. This is further investigated in this section. The investigation involves examining how sensitive the log likelihood function is to a change in the parameter π_i and which values of π_i in the set $0 < \pi_i < 1$ result in significant autocorrelation in the squared residuals ϵ_{it}^2.

In the context of fitting an O-EWMA model, the criterion for determining whether a series is considered to display significant autocorrelation or not needs to be discussed. In this context, a series is considered to display significant autocorrelation if the p-value, associated with the Q-statistic fitted to the series, is smaller than 0.05. This Q-statistic is fitted under the null hypothesis that the each of the lags up to and including the number of lags tested are zero. Therefore the p-value is the probability of observing a larger Q-statistic. A Q-statistic at lags 1, 5 and 10 are calculated for each series of squared residuals ϵ_{it}^2 for $i = 1, 2, \cdots N$. Based on the findings, one of the three lags will be selected to be used for the remainder of this study when fitting an O-EWMA model.

Consequently, for each series of squared residuals it is possible to determine a subset of the set $0 < \pi_i < 1$ such that none of the values of π_i in this subset induced significant autocorrelation in the squared residuals. In other words the p-value, associated with the Q-statistic fitted to that series of squared residuals, is greater than 0.05. The parameter π_i is estimated by maximising the likelihood function over this subset of values. However, if no values of π_i in the set $0 < \pi_i < 1$ induce significant autocorrelation then the estimate of π_i is what is typically referred to as the maximum likelihood estimate of an IGARCH(1,1) model. In such a case, this estimate of π_i will be referred to as the unrestricted maximum likelihood estimate. On the other hand, if there are values of π_i in the set $0 < \pi_i < 1$ which induce significant autocorrelation then the resulting estimate of π_i is the value which maximises the likelihood function over this smaller subset. In such a case, this estimate of π_i will be referred to as the restricted maximum likelihood estimate. However, if all values of π_i between 0 and 1 induce significant autocorrelation in the squared residuals then the unrestricted maximum likelihood estimate is used for π_i.

Since the estimate of π_i may be the restricted maximum likelihood estimate and not the unrestricted maximum likelihood estimate, it is important to consider whether the value of the log likelihood at the restricted maximum likelihood estimate of π_i is close to the maximum of the log likelihood function. Note that the lag used in the Q-statistic will affect the estimate of π_i and therefore the model results.

A plot of the log likelihood of each series of standardised principal component scores can be found in Figure 7.4 for the share data set and in Figure 7.5 for the exchange rate data set. The red dots on each series indicate where the log likelihood is a maximum over the range $0 < \pi_i < 1$ and the associated unrestricted maximum likelihood estimate of the parameter π_i. The value of the unrestricted maximum likelihood estimates of π_i are also indicated in the legend. As the shape of the log likelihood is very similar regardless of the choice of conditional mean model, $\boldsymbol{B}$ and $\boldsymbol{W}$ only one plot is included for each data set. In fact, either choice of $\boldsymbol{W}$ give identical results for the O-EWMA models so the choice of $\boldsymbol{W}$ is irrelevant. In these plots the conditional mean is modelled using an ARMA model, $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ and $\boldsymbol{W} = \boldsymbol{I}$.

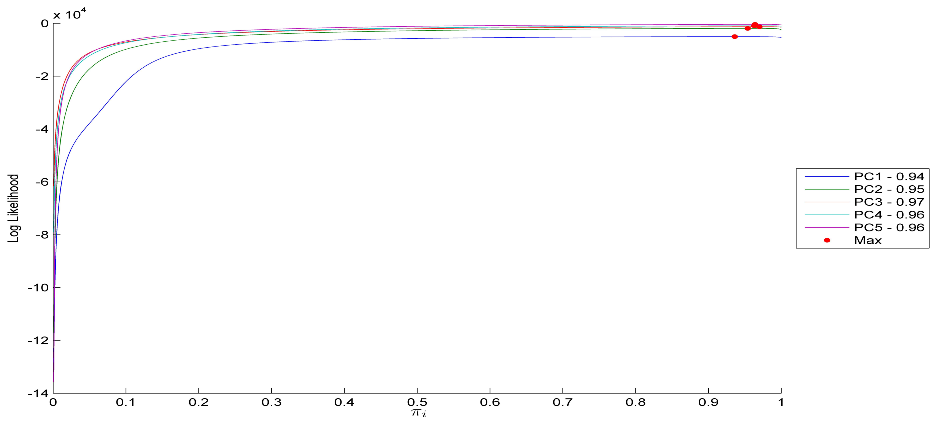


Figure 7.4: Shares: Plot of the Log Likelihood of the O-EWMA Model

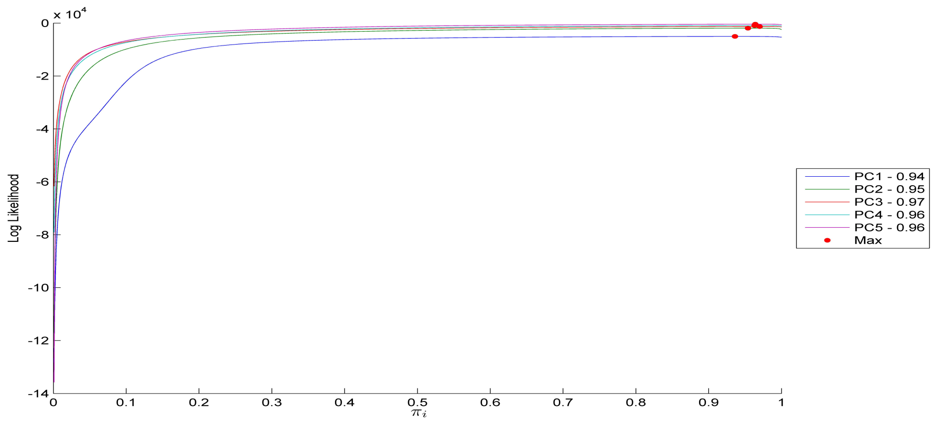


Figure 7.5: Exchange Rates: Plot of the Log Likelihood of the O-EWMA Model

For both data sets it is evident that choosing π_i anywhere between 0.3 and 1 result in the log likelihood function having a value which is similar to the maximum value of the log likelihood over the range $0 < \pi_i < 1$. Hence if the restricted maximum likelihood estimate is between 0.3 and 1 it should not have a big negative impact on the likelihood function.

The p-values, associated with Q(1), Q(5) and Q(10) fitted to the squared residuals, are plotted for each series for both data sets. The plots for the share data set are found in Figures 7.6, 7.7 and 7.8 and for the exchange rate data set in Figures 7.9, 7.10 and 7.11. The red dot on each series represent the restricted maximum likelihood estimate of π_i and the value of these estimates are also given on the legend. Therefore for any one series, the likelihood is maximised over the subset of values of π_i between 0 and 1 where the associated p-values are above the horizontal black line drawn at 0.05. For example, the first series PC1 in Figure 7.6 will estimate π_i by maximising the likelihood over the subset $0 < \pi_i < 0.86$ as this is the subset over which the p-values are greater than 0.05.

In these plots the conditional mean is modelled as an ARMA process, $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ and $\boldsymbol{W} = \boldsymbol{I}$. However, the plots of the p-values do vary somewhat across the different choices of conditional mean model and $\boldsymbol{B}$ but the general idea is the same. They are similar in the sense that for most series the p-values are the largest for values of π_i close to 0 and values close to 1 but are very small in between the two. This is evident in the plots for both the share data set and the exchange rate data set.

Hence it is apparent from the plots of the p-values that in general the restricted maximum likelihood estimate of π_i is between 0.3 and 1 so the log likelihood at this value of π_i is similar to the maximum value of the log likelihood. Furthermore the difference between the unrestricted and restricted maximum likelihood estimates of π_i, for any one series, are much greater for the share data set than the exchange rate data set and, in fact there is hardly any difference in the estimates in the case of the exchange rate data set.

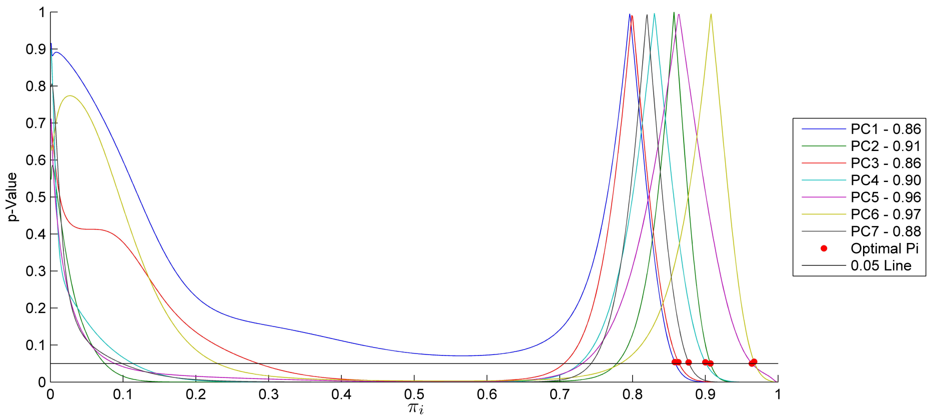


Figure 7.6: Shares: Plot of the p-values of Q(1) fitted to the squared residuals of the O-EWMA Model

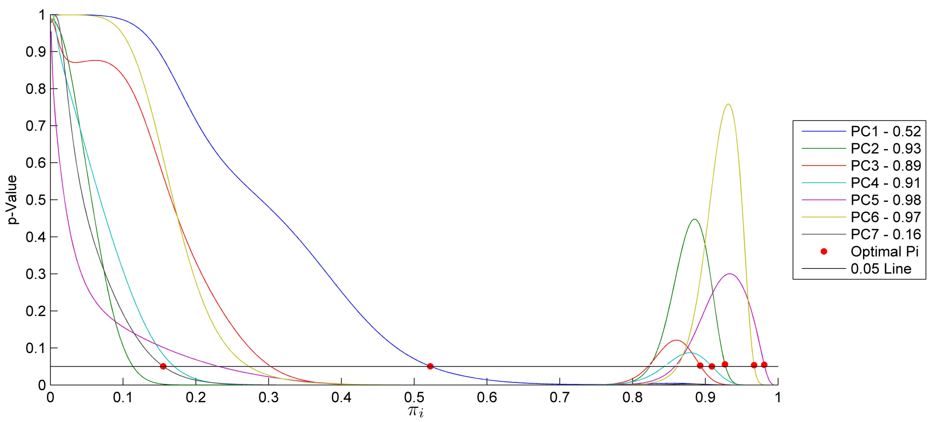


Figure 7.7: Shares: Plot of the p-values of Q(5) fitted to the squared residuals of the O-EWMA Model

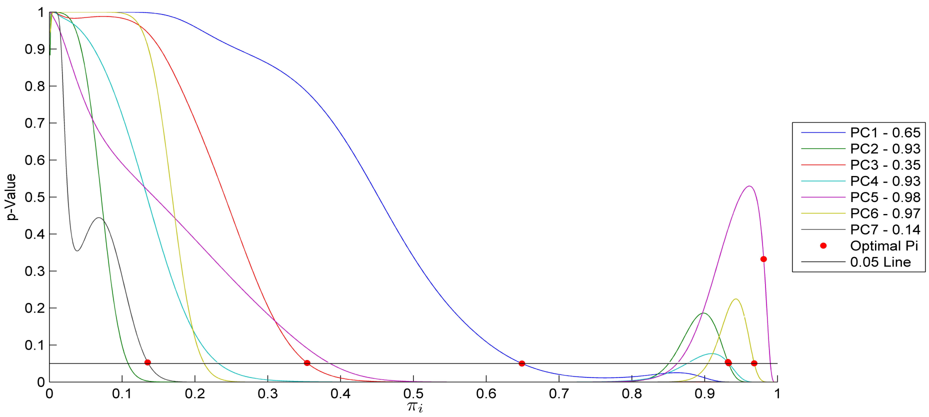


Figure 7.8: Shares: Plot of the p-values of Q(10) fitted to the squared residuals of the O-EWMA Model

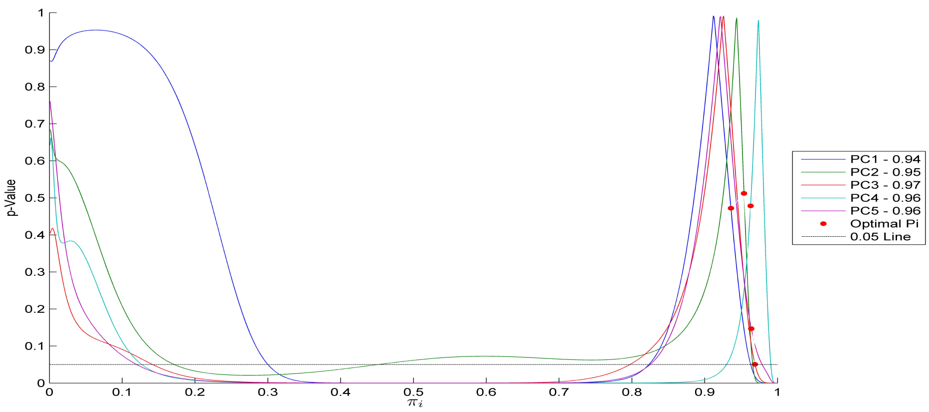


Figure 7.9: Exchange Rates: Plot of the p-values of Q(1) fitted to the squared residuals of the O-EWMA Model

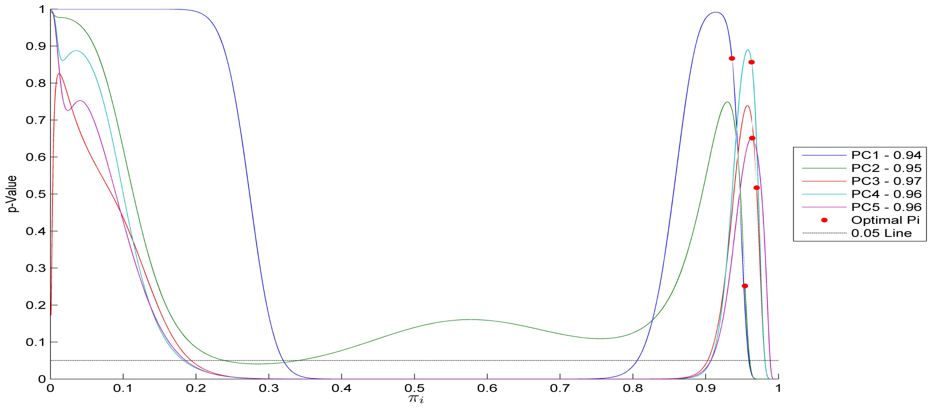


Figure 7.10: Exchange Rates: Plot of the p-values of Q(5) fitted to the squared residuals of the O-EWMA Model

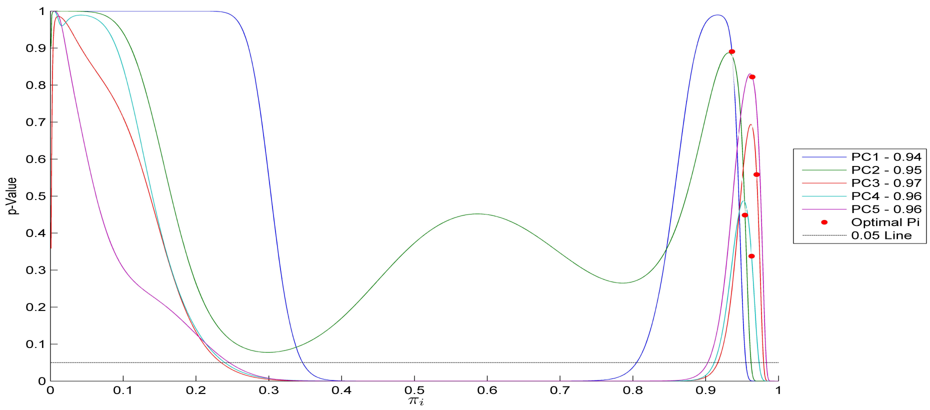


Figure 7.11: Exchange Rates: Plot of the p-values of Q(10) fitted to the squared residuals of the O-EWMA Model

For completeness, tables of the unrestricted and restricted maximum likelihood estimates of π_i for all the variations of the conditional mean model and $\boldsymbol{B}$ are included. These estimates of π_i for the share data set can be found in Tables 7.9, 7.10, 7.11 and 7.12. From the tables, it is evident that for a given series, conditional mean model and choice of $\boldsymbol{B}$ the unrestricted estimate of π_i is usually greater than any one of the restricted estimates of π_i. Hence the persistence of the conditional volatility is greatest when the unrestricted estimate is used. The restricted estimates are usually smaller because as the number of lags used in the Q-statistic increase, at least up to lag 10, so the p-values tend to get closer to zero at values of π_i close to 1. Moreover, for a given series, the unrestricted maximum likelihood estimates are very similar across the different conditional mean models and $\boldsymbol{B}$ but where the restricted maximum likelihood estimates are concerned they start to differ across the variations. This is mostly the case in Tables 7.11 and 7.12.

On the other hand the unrestricted and restricted maximum likelihood estimates of π_i for the exchange rate data set can be found in Tables 7.13, 7.14, 7.15 and 7.16. Unlike the estimates of π_i for the share data, the unrestricted maximum likelihood estimates of π_i for any one series is very similar to any one of the three restricted maximum likelihood estimates of π_i. Additionally the estimates of π_i in any one of the tables for a given series are very similar across the different choices of conditional mean model and $\boldsymbol{B}$.

It is interesting that the unrestricted maximum likelihood estimates of π_i for both data sets, that is in Tables 7.9 and 7.13, are generally slightly greater than the comparable O-GARCH estimates of β_i. Therefore the conditional volatility of the O-EWMA models are slightly more persistent and less responsive to market movements than the O-GARCH models. Moreover, according to J.P. Morgan and Reuters (1996) a value of 0.94 is generally what one would expect to be best for an EMWA model of the squared or cross product of the returns which is more or less the region of the unrestricted maximum likelihood estimates of π_i in Tables 7.9 and 7.13.

Table 7.9: Shares - Unrestricted Maximum Likelihood estimate of the O-EWMA parameter π_i

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$	$\boldsymbol{B}$	Number of the Principal Component						
		1	2	3	4	5	6	7
ARMA	$\boldsymbol{I}$	0.9410	0.9540	0.9800	0.9360	0.9810	0.9720	0.9620
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.9380	0.9560	0.9790	0.9440	0.9810	0.9720	0.9620
VAR	$\boldsymbol{I}$	0.9390	0.9540	0.9810	0.9340	0.9820	0.9720	0.9640
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.9400	0.9570	0.9780	0.9440	0.9820	0.9720	0.9640

Table 7.10: Shares - Restricted Maximum Likelihood estimate of the O-EWMA parameter π_i using Q(1) to determine the subset of possible parameters

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$	$\boldsymbol{B}$	Number of the Principal Component						
		1	2	3	4	5	6	7
ARMA	$\boldsymbol{I}$	0.8540	0.8980	0.8630	0.8980	0.9610	0.9690	0.8770
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.8580	0.9070	0.8630	0.9000	0.9640	0.9670	0.8770
VAR	$\boldsymbol{I}$	0.8600	0.8950	0.8540	0.8930	0.9790	0.9720	0.9040
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.8680	0.9050	0.8640	0.8920	0.9810	0.9720	0.9050

Table 7.11: Shares - Restricted Maximum Likelihood estimate of the O-EWMA parameter π_i using Q(5) to determine the subset of possible parameters

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$	$\boldsymbol{B}$	Number of the Principal Component						
		1	2	3	4	5	6	7
ARMA	$\boldsymbol{I}$	0.4250	0.9230	0.8980	0.9110	0.9810	0.9680	0.1180
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.5220	0.9270	0.8930	0.9090	0.9810	0.9670	0.1550
VAR	$\boldsymbol{I}$	0.5380	0.9230	0.8770	0.9060	0.9810	0.9720	0.3210
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.6770	0.9290	0.8840	0.9110	0.9820	0.9720	0.3040

Table 7.12: Shares - Restricted Maximum Likelihood estimate of the O-EWMA parameter π_i using Q(10) to determine the subset of possible parameters

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$	$\boldsymbol{B}$	Number of the Principal Component						
		1	2	3	4	5	6	7
ARMA	$\boldsymbol{I}$	0.5010	0.9380	0.4020	0.9300	0.9810	0.9680	0.1440
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.6490	0.9320	0.3540	0.9330	0.9810	0.9680	0.1350
VAR	$\boldsymbol{I}$	0.6350	0.9350	0.3270	0.9220	0.9820	0.9720	0.3550
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.7710	0.9310	0.3000	0.9340	0.9820	0.9720	0.3370

Table 7.13: Exchange Rates - Unrestricted Maximum Likelihood estimate of the O-EWMA parameter π_i

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$	$\boldsymbol{B}$	Number of the Principal Component				
		1	2	3	4	5
ARMA	$\boldsymbol{I}$	0.9350	0.9560	0.9660	0.9700	0.9640
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.9360	0.9540	0.9700	0.9630	0.9640
VAR	$\boldsymbol{I}$	0.9390	0.9560	0.9620	0.9720	0.9630
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.9400	0.9540	0.9650	0.9650	0.9650

Table 7.14: Exchange Rates - Restricted Maximum Likelihood estimate of the O-EWMA parameter π_i using Q(1) to determine the subset of possible parameters

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	Number of the Principal Component				
		1	2	3	4	5
ARMA	$\boldsymbol{I}$	0.9350	0.9560	0.9590	0.9700	0.9640
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.9360	0.9540	0.9690	0.9630	0.9640
VAR	$\boldsymbol{I}$	0.9390	0.9540	0.9540	0.9720	0.9630
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.9400	0.9540	0.9650	0.9650	0.9650

Table 7.15: Exchange Rates - Restricted Maximum Likelihood estimate of the O-EWMA parameter π_i using Q(5) to determine the subset of possible parameters

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	Number of the Principal Component				
		1	2	3	4	5
ARMA	$\boldsymbol{I}$	0.9350	0.9560	0.9660	0.9700	0.9640
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.9360	0.9540	0.9700	0.9630	0.9640
VAR	$\boldsymbol{I}$	0.9390	0.9540	0.9620	0.9720	0.9630
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.9400	0.9540	0.9650	0.9650	0.9650

Table 7.16: Exchange Rates - Restricted Maximum Likelihood estimate of the O-EWMA parameter π_i using Q(10) to determine the subset of possible parameters

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	Number of the Principal Component				
		1	2	3	4	5
ARMA	$\boldsymbol{I}$	0.9350	0.9560	0.9660	0.9700	0.9640
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.9360	0.9540	0.9700	0.9630	0.9640
VAR	$\boldsymbol{I}$	0.9390	0.9560	0.9620	0.9720	0.9630
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	0.9400	0.9540	0.9650	0.9650	0.9650

However the value of π_i used in any O-EWMA model from this point onwards will be the restricted maximum likelihood estimate where the relevant Q-statistic tests autocorrelations up to and including lag 5. Hence the likelihood is maximised over the subset of $0 < \pi_i < 1$ where the p-value of Q(5), fitted to the squared residuals under the null hypothesis that all autocorrelations up to and including lag 5 are zero, is greater than 0.05. Recall that this choice was mentioned in Section 3.5.1. The reasoning behind this is that a lag of 5 will ensure that any weekly effects are not present and the lag chosen is small enough to ensure that the test still has sufficient power. Therefore using a lag of 5 is a sensible choice to make.

The O-GARCH estimates of β_i should in fact be compared to these restricted maximum likelihood estimates of π_i since these are the values which will actually be used in the O-EWMA models fitted in this study. However, this makes virtually no difference to the exchange rate data set but it does affect the comparison in the case of the share data set. In the case of the share data the restricted estimates of π_i in Table 7.11, that is those selected using a Q-statistic at lag 5, are considerably smaller than the O-GARCH estimates of β_i for series 1 and 7 and are therefore more responsive to market movements. However for the remaining series they are still generally larger.

Note that unlike the GARCH(1,1) models, a t-statistic for the estimates of π_i and the p-values associated with the t-statistics are not calculated for the IGARCH(1,1) models. The reason for this is that the set of π_i over which the likelihood is maximised is determined by the autocorrelation of the resulting squared residuals. Therefore it may be difficult to determine an appropriate p-value, associated with the t-statistic, for the null hypothesis that π_i is zero.

7.2.3 O-SV Parameter Estimates

The final conditional volatility model fitted to the standardised principal component scores is the stochastic volatility model of Shephard (1994). Recall that the main equation of this model is

$$\gamma_{it} \quad = \quad exp(r_{it})\, \gamma_{i,t-1}\, \eta_{it}$$

where η_{it} follows a beta distribution such that

$$\eta_{it} \quad \sim \quad beta(\omega_i\, a_{i,t-1},\, (1-\omega_i)\, a_{i,t-1}).$$

Since the precision γ_{it} is a random variable and not a constant, a point estimate for the conditional volatility is required. The point estimate used for the conditional volatility of the i^{th} series at time t is $1/E[\gamma_{it}|\gamma_{i,t-1}]$, as discussed in Chapter 3.

This section provides some justification for the fact that the log likelihood of the conditional variance of the i^{th} series of standardised principal components is only maximised with respect to ω_i and the values of the starting parameters a_{i0} and b_{i0} are kept fixed. To investigate this the log likelihood is maximised with respect to ω_i over an appropriate grid of values of a_{i0} and b_{i0}. Thus by varying the values of a_{i0} and b_{i0}, a plot of the maximum likelihood estimate of ω_i against the values a_{i0} and b_{i0} can be constructed. Furthermore the maximum value of the log likelihood, with respect to ω_i, can be plotted against a grid of fixed values of the parameters a_{i0} and b_{i0}. Together these plots give an indication of the effect the starting parameter a_{i0} and b_{i0} have on the results. These plots are similar for any one of the choices of conditional mean model, $\boldsymbol{B}$ and $\boldsymbol{W}$ so for illustrative purposes a choice is made for each and used throughout the plots presented in this section.

Firstly the share data set is considered. Figure 7.12 contains a plot of the maximum value of the log likelihood of the first series, with respect to ω_1, against a grid of values of a_{10} and b_{10}. The shape of the curve is similar for the remaining six series of standardised principal component scores so it is not necessary to include these plots. Therefore from the plot it is evident that the maximum value of the log likelihood is similar for most values of the parameters a_{10} and b_{10} on the grid. The only values which give slightly poorer results are when a low value of b_{10} is simultaneously chosen with a high value of a_{10}. In fact choosing the starting parameters a_{10} and b_{10} close to zero but not exactly zero result in one of the better outcomes. The reasons that they should not be exactly zero are discussed in Section 3.6.

Additionally a plot of the value of ω_1 which maximises the log likelihood of the first series for various values of a_{10} and b_{10} is given in Figure 7.13. The shape of the curve is similar to that of the maximum log likelihood curve and the range of values of ω_1 on the plot are very small. The shape of the curve is similar for each of the remaining six series and the range of ω_i on each of these other curves is also small.

Thus for the share data, the results should be similar regardless of the choice of starting parameters a_{i0} and b_{i0}. However a low value of b_{i0} and a high value of a_{i0} will give slightly less optimal results and should be avoided.

The exchange rate data set is now considered. Figure 7.14 and 7.15 contain

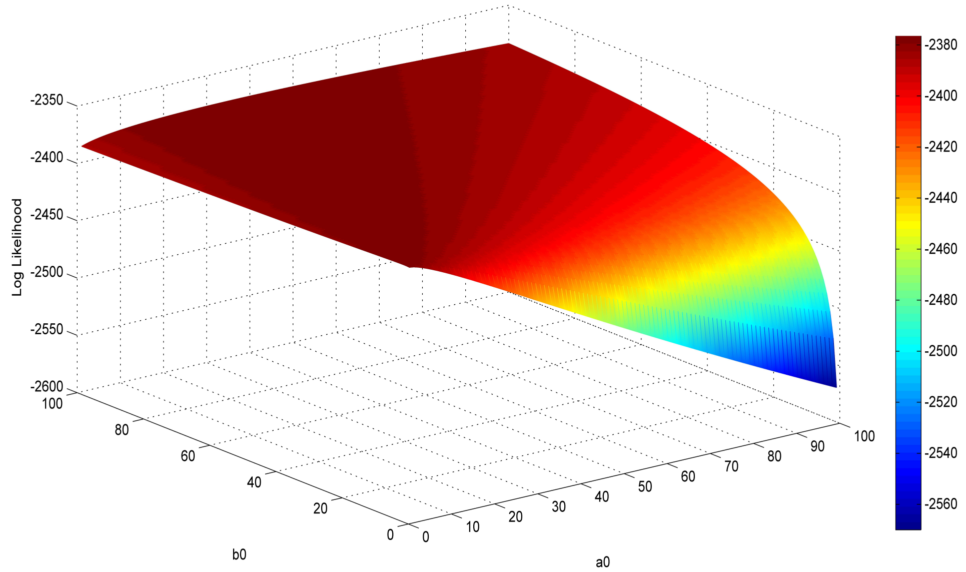


Figure 7.12: Shares: Maximum of the Log Likelihood $LL(u_{11}, \cdots, u_{1T}, \omega_1, a_{10}, b_{10})$ for given values of a_{10} and b_{10}

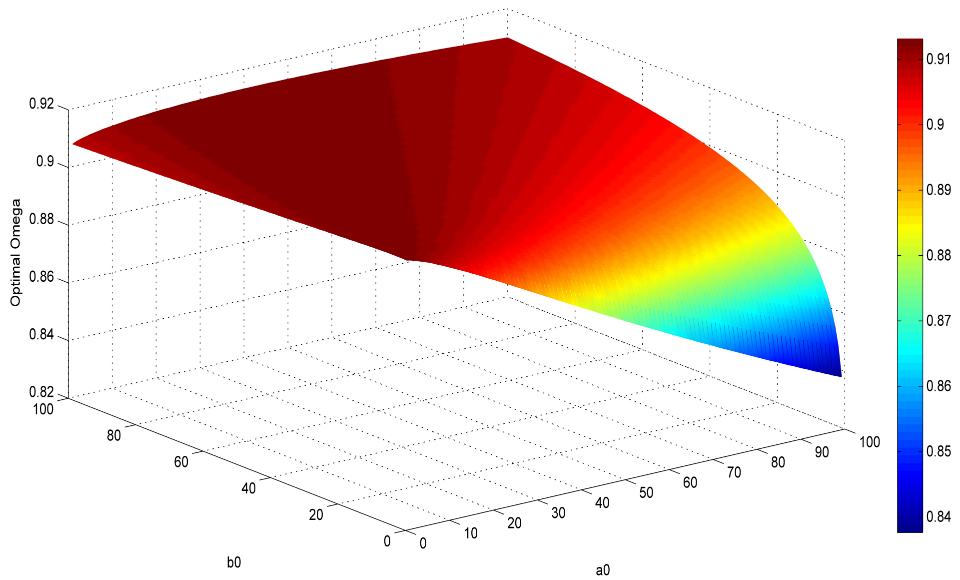


Figure 7.13: Shares: Maximum Likelihood estimates of ω_1 for given values of a_{10} and b_{10}

a plot of the maximum value of the log likelihood for various values of a_{i0} and b_{i0} for the first and fifth series respectively. Hence it is evident from the two figures that the shape of the curves are quite different. The reason for choosing these two series in particular is that as the number of the series gets closer to five so the shape of the curve looks more like Figure 7.15 and less like Figure 7.14 but as the number of the series gets closer to one so the curve looks more like Figure 7.14 and less like Figure 7.15. In other words, the curve of any series between one and five look like a combination of these two curves.

Although the shape of the curves vary for each series, they all have a relatively small range of log likelihood values over the grid of a_{i0} and b_{i0}. Thus the choice of a_{i0} and b_{i0} do not have a big impact on the value of the log likelihood. In both these figures choosing a_{i0} and b_{i0} close to zero but not exactly zero, due to the difficulties mentioned in Section 3.6, result in the maximum log likelihood value being one of the larger values on their respective figures.

Furthermore a plot of the values of ω_1 which maximise the log likelihood of the first series is given in Figure 7.16 and a plot of the values of ω_5 which maximise the log likelihood of the fifth series is given in Figure 7.17. Once again, the shape of the curve for any series between one and five is somewhere in-between the shapes of the curves in Figures 7.16 and 7.17. Additionally the shape of each of these curves are similar to that of their respective maximum log likelihood curves. However, the range of values of ω_i on any one of the curves for the five series is small so the starting parameters a_{i0} and b_{i0} make little difference to the results. Thus these plots once again demonstrate that for the exchange rate data the choice of the starting parameters a_{i0} and b_{i0} only have a small effect on the results.

Hence in both data sets choosing a_{i0} and b_{i0} close to zero but not exactly zero for each series results in one of the better outcomes within each series. In addition the parameters a_{i0} and b_{i0} have very little impact on the results.

Consequently the parameters a_{i0} and b_{i0} are fixed for two reasons. The first is that the optimisation algorithms in MATLAB have difficulty converging when the log likelihood is maximised with respect to all three parameters ω_i, a_{i0} and b_{i0} simultaneously. The second is that the choice of the starting parameters a_{i0} and b_{i0} have little impact on the log likelihood and the maximum likelihood estimate of the parameter ω_i. As the choice of a_{i0} and b_{i0} have little effect on the results, it is not necessary to invoke an optimiser which can handle such a problem.

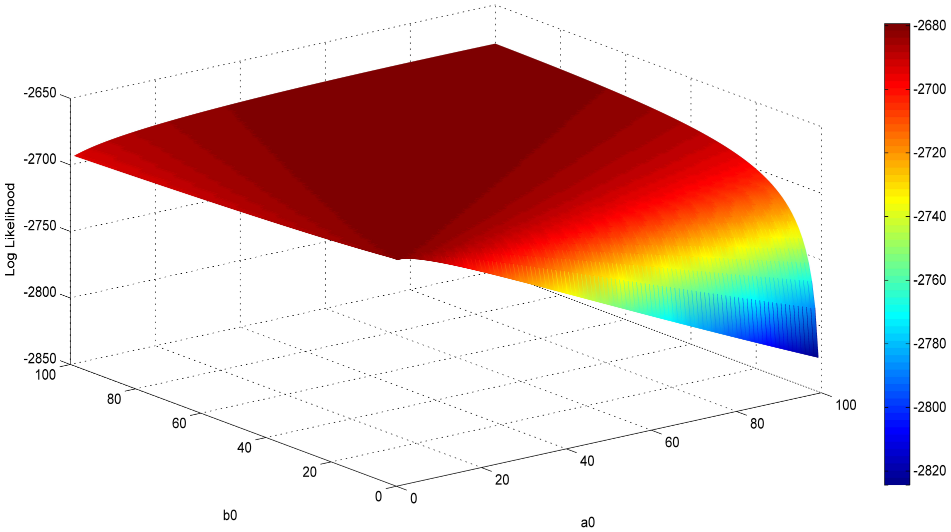


Figure 7.14: Exchange Rates: Maximum of the Log Likelihood $LL(u_{11}, \cdots, u_{1T}, \omega_1, a_{10}, b_{10})$ for given values of a_{10} and b_{10}

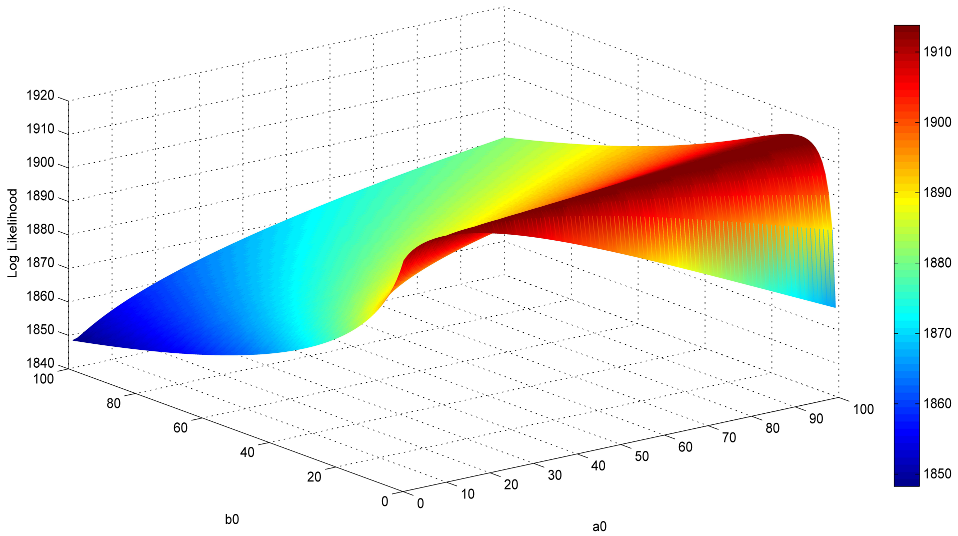


Figure 7.15: Exchange Rates: Maximum of the Log Likelihood $LL(u_{51}, \cdots, u_{5T}, \omega_5, a_{50}, b_{50})$ for given values of a_{50} and b_{50}

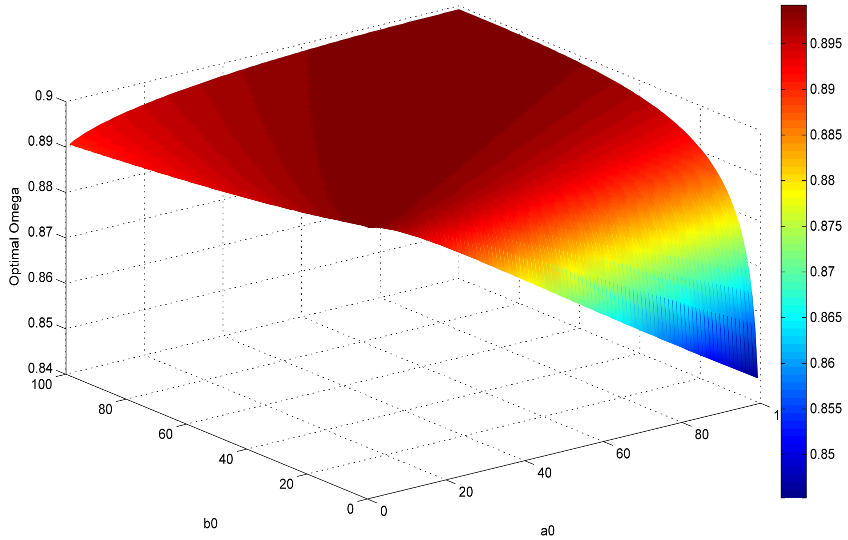


Figure 7.16: Exchange Rates: Maximum Likelihood estimates of ω_1 for given values of a_{10} and b_{10}

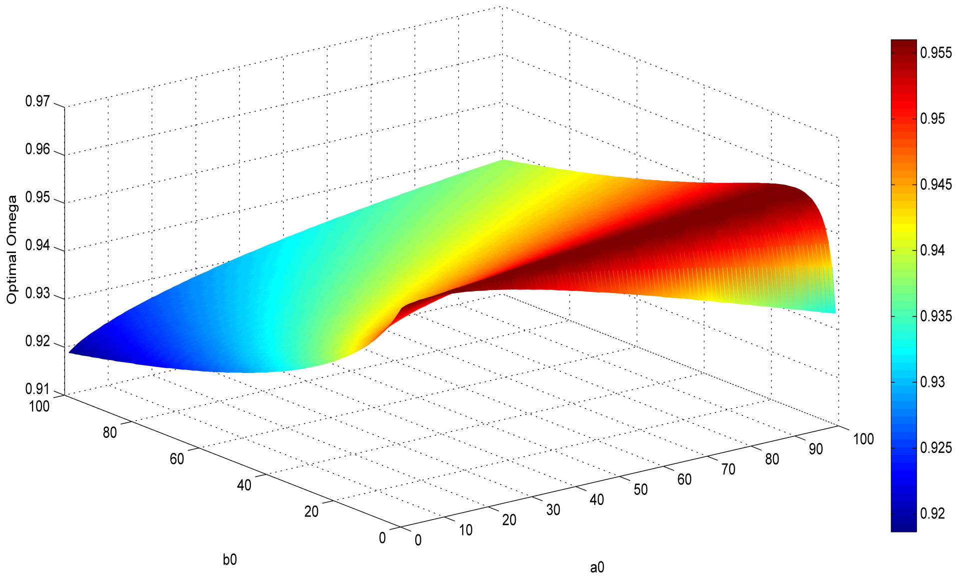


Figure 7.17: Exchange Rates: Maximum Likelihood estimates of ω_5 for given values of a_{50} and b_{50}

Since the justification for the method of parameter estimation has been discussed the actual values of ω_i are now considered. These parameter estimates are found in Tables 7.17 and 7.18.

It is evident from these tables that the values of ω_i are fairly large which implies that the precision γ_{it} changes slowly. Therefore the conditional volatility can be said to be fairly persistent. Another observation which can be made is that within any one series, the estimates of ω_i are similar across the various choices of conditional mean model, $\boldsymbol{B}$ and $\boldsymbol{W}$. In fact in the case where $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ the choice of $\boldsymbol{W}$ results in identical estimates of ω_i up to the third decimal place. In the case of $\boldsymbol{B} = \boldsymbol{I}$ the choice of $\boldsymbol{W}$ still makes little difference but the difference is typically greater than in the case of $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$. A similar phenomenon is observed in the estimates of the O-GARCH parameters, that is when $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ either choice of $\boldsymbol{W}$ give extremely similar parameter estimates. Thus it appears that when $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$, the suggestion made by Bongers (2008) is of little use and in fact, is irrelevant in the case of the O-EWMA model.

Recall that the conditional variance estimate of the O-SV model is similar to that of the O-EWMA model and that ω_i is comparable to π_i under certain conditions. Similarly the O-SV parameter is also comparable to the O-GARCH parameter β_i. Although all three of these parameters measure persistence they do so in different ways which makes strict comparison difficult. However what can be concluded is that all three models, O-GARCH, O-EWMA and O-SV, suggest that the conditional volatility is fairly persistent.

Table 7.17: Shares - SV parameter estimate of ω_i

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	Number of the Principal Component						
			1	2	3	4	5	6	7
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	0.9265	0.9335	0.9400	0.9136	0.9595	0.9607	0.9604
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.9260	0.9331	0.9403	0.9132	0.9674	0.9665	0.9645
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.9208	0.9426	0.9512	0.9265	0.9671	0.9663	0.9645
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.9209	0.9426	0.9512	0.9265	0.9671	0.9663	0.9644
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	0.9295	0.9336	0.9475	0.9088	0.9606	0.9629	0.9624
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.9292	0.9332	0.9495	0.9081	0.9693	0.9682	0.9675
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.9238	0.9438	0.9549	0.9255	0.9692	0.9683	0.9674
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.9239	0.9438	0.9549	0.9255	0.9692	0.9682	0.9673

Table 7.18: Exchange Rates - SV parameter estimate of ω_i

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	Number of the Principal Component				
			1	2	3	4	5
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	0.9073	0.9367	0.9467	0.9490	0.9452
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.9093	0.9396	0.9562	0.9630	0.9569
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.9085	0.9412	0.9645	0.9584	0.9589
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.9086	0.9412	0.9645	0.9584	0.9588
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	0.9114	0.9387	0.9444	0.9470	0.9439
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.9143	0.9422	0.9519	0.9613	0.9551
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	0.9131	0.9434	0.9594	0.9570	0.9591
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	0.9131	0.9434	0.9594	0.9570	0.9591

Chapter 8

Results - Analysis of O-GARCH, O-EWMA and O-SV

There are a number of methods which can be used to evaluate the fit of a model and the accuracy of its forecasts. Two methods are used to analyse the results. These are by

1. Graphical output which is used to get a general overview of the results before performing a more detailed analysis.

2. Various statistics which are used to give an objective evaluation of the models considered.

Another possible method which could be used is simulation. However the models in this thesis are not evaluated using simulation for two reasons. The first is the flawed assumption that the form of the model generating share prices and exchange rates is known. This is particularly problematic when modelling share and exchange rate returns as it is unlikely that a parsimonious models exist because the values are determined by market consensus, in other words an average of the opinions of all market participants. These opinions are formed by thousands of different pieces of information and the interpretation of how these will affect share prices and exchange rates. Therefore if a model fits the simulated data well it does not imply that the model will fit real world data well as the simulated data might not be a good representation of real world data.

The second reason is a problem specific to simulating from an orthogonal factor model. The problem is that there may be more than one orthogonal matrix $\boldsymbol{A}$ and vector $\boldsymbol{x_t}$ that can generate the same adjusted residuals $\boldsymbol{y_t}$. Therefore even if the model fits the simulated data well, the model estimates of $\boldsymbol{A}$ and $\boldsymbol{x_t}$ may not be close to those that were used to simulate the data.

Thus in short this chapter examines the model results via graphical output and various statistics. In terms of graphical output it contains plots of the model estimates of the conditional variances and correlations which are compared to plots of the underlying data and in terms of statistics three different statistics are calculated using the model results.

8.1 Proxy for the True Conditional Covariance

There are some difficulties in this thesis with regards to evaluating the models because in order to evaluate a model the true value of what is being modelled should ideally be known. In this case the conditional covariances of the log returns $\boldsymbol{z_t}$ are being modelled but the true value of these conditional covariances are unknown. Therefore a proxy is required for the conditional covariances.

It is intuitive to proxy the conditional covariance of the i^{th} and j^{th} series of the log returns $\boldsymbol{z_t}$ by the cross product of the i^{th} and j^{th} series of the conditional mean model residuals $\boldsymbol{\hat{e}_t} = \boldsymbol{z_t} - \boldsymbol{\hat{\mu}_t}$. Hence this approximation can be represented as

$$\begin{aligned} Cov[z_{it}\, z_{jt}|F_{t-1}] &\approx (z_{it} - \hat{\mu}_{it})(z_{jt} - \hat{\mu}_{jt}) \\ &= \hat{e}_{it}\hat{e}_{jt}. \end{aligned}$$

This proxy can in fact be represented more generally. The more general proxy introduced here is the proxy which is implied in the AMAD (adaptive mean absolute deviations) statistic in Fan et al. (2008). Therefore a more general proxy for the conditional covariance of the i^{th} and j^{th} series of the log returns $\boldsymbol{z_t}$ is the sum of the cross products the i^{th} and j^{th} series of the conditional mean model residuals in the vicinity of time t so that

$$\begin{aligned} Cov[z_{it}\, z_{jt}|F_{t-1}] &\approx \sum_{l=-v}^{l=v} (z_{i,t+l} - \hat{\mu}_{i,t+l})(z_{j,t+l} - \hat{\mu}_{j,t+l}) \\ &= \sum_{l=-v}^{l=v} \hat{e}_{i,t+l}\hat{e}_{j,t+l}. \end{aligned}$$

The sum is from time $t-v$ to time $t+v$ where v is a relatively small integer. Hence if $v=0$ then this proxy is simply $\hat{e}_{it}\hat{e}_{jt}$. On the other hand if v is not zero then including adjacent residuals may help to obtain a better estimate by removing some of the background noise (Fan et al., 2008). However the approach could have the disadvantage of giving a less accurate point estimate of the conditional covariance as the estimate essentially smooth the series of the conditional covariances.

Since a suitable proxy for the conditional covariances has been introduced the actual results can now be compared to this proxy to evaluate model fit and the accuracy of the forecasts.

8.2 Visual Comparison of Results

Before evaluating the models by means of a range of statistics measuring model fit, it is beneficial to get an overview of the results first. This is important because statistics can lose valuable information as they summarise the information into one number. Therefore to get an overview of the results a few representative plots for the various models are presented. Although all the plots are not included those that are representative are shown and give some idea of how reasonable the estimates are.

Recall that each of the O-GARCH, O-EWMA and O-SV models have two choices for the conditional mean model, and for $\boldsymbol{B}$ and $\boldsymbol{W}$. However the choice of conditional mean model, $\boldsymbol{B}$ and $\boldsymbol{W}$ result in very similar plots within each of the O-GARCH, O-EWMA and O-SV models. Therefore it is only necessary to plot one combination of these variations and thus an ARMA model, $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ and $\boldsymbol{W} = \boldsymbol{I}$ is selected for fitting the three models.

However including only one set of the variations for each of the three models still results in a large number of plots. This is because the shares returns have 28 different series of conditional variances and covariances and the exchange rate returns have 15 different series of conditional covariances. Thus to prevent the number of plots included from becoming too great the conditional variance of only one series of returns and the conditional correlation of only one pair of returns are plotted for each data set.

Additionally plots of the underlying data are also presented in order to ascertain whether these estimates are reasonable. Firstly the log returns of a series are plotted next to the conditional variance estimates of that series to get a sense of which periods should have high volatility and which should have low volatility. Secondly the cross product of a pair of returns are plotted next to the conditional correlation estimates of that pair of returns. However it is difficult to compare the plots of the cross product of the returns to the conditional correlation estimates. This is because time periods where the cross product appears to be more volatile may be due to one or both of the return series being more volatile and not due to increased correlation. The conditional covariance estimates could have been plotted instead of the correlations as this is easier to compare to the cross product. However the reason the correlations are plotted is that it is easier to determine whether a correlation estimate is reasonable than whether a covariance estimate is reasonable.

In the case of the share data the first set of plots are of the model estimates of the conditional variances of ABSA's returns shown in Figure 8.1. Note that the range of the y-axis for the O-GARCH and O-SV models are the same but that of the O-EWMA model is about twice the range of the other two. These conditional variances are daily estimates and have not been annualised. The log returns of ABSA are also included in Figure 8.1 to get a sense of how reasonable the estimates are. In this figure it can be observed that the majority of the conditional variance estimates are in the region of 0.472×10^{-3} which is the unconditional daily sample variance of ABSA's returns over the period. Therefore the level of the estimates appear to be reasonable. Moreover all three models appear to capture the volatility in periods where there is increased volatility. One example of this can be seen from May 2007 to May 2009, where the sub prime crisis resulted in volatile financial market. Furthermore if the models are compared then it appears that the variance estimates of the O-EWMA model are the most responsive to the observed share returns and O-SV the least responsive.

The second set of plots for the share data set are of the model estimates of the conditional correlations of ABSA and FirstRand's returns and are shown in Figure 8.2. This figure also contains a plot of the cross product of ABSA and FirstRand's returns. However it is difficult to determine whether these correlation estimates are appropriate. The estimates do not seem unreasonable with regards to the average level of the estimates in light of the fact that the unconditional sample correlation of ABSA and FirstRand's returns over the period is 0.59 (see Table 5.2). On the other hand the conditional correlation estimates appear to be fairly volatile, especially those of the O-

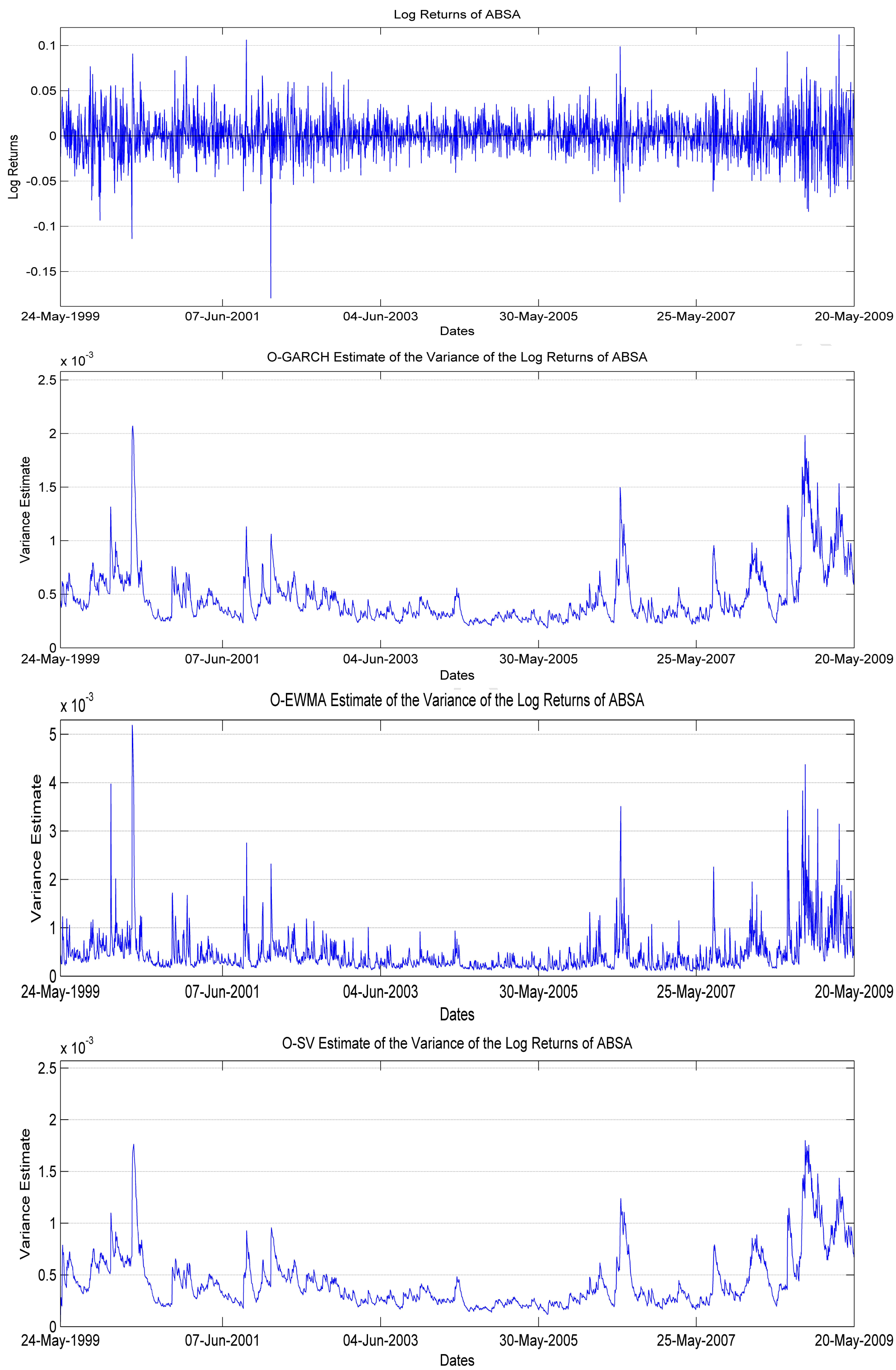


Figure 8.1: Shares: Log Returns of ABSA and the Variance Estimates of these Returns for the O-GARCH, O-EWMA and O-SV models

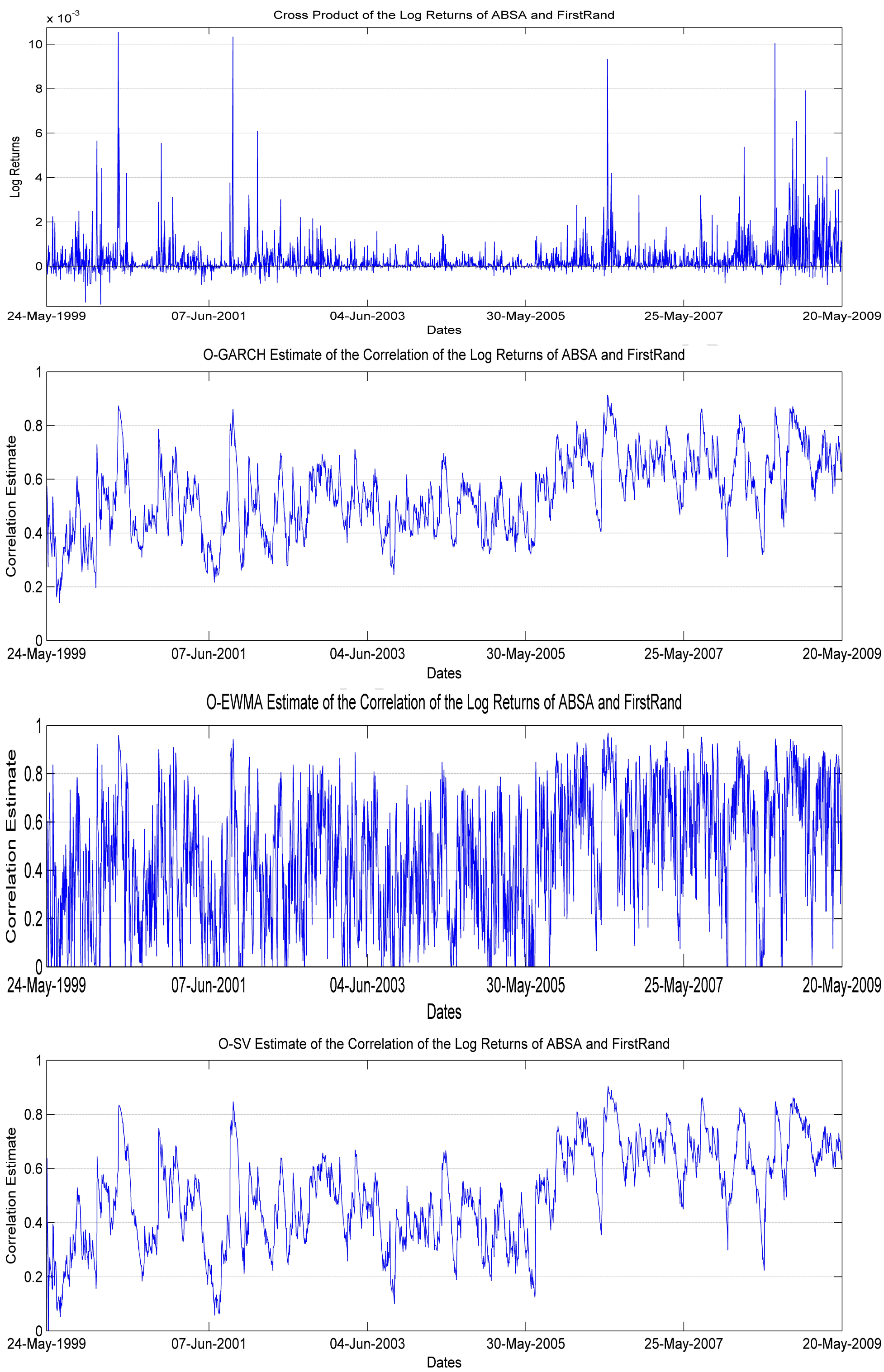


Figure 8.2: Shares: Cross Product and Correlation Estimates of the Log Returns of ABSA and FirstRand for the O-GARCH, O-EWMA and O-SV models

EWMA model, so the question must be asked as to whether one expects the correlations to be so volatile. Typically the returns of two banks operating in similar geographical regions are usually expected to be fairly correlated so it is questionable as to whether the conditional correlation estimates should be so volatile and whether some of the conditional correlations should ever be so close to zero. With regards to volatility of the estimates, the O-EWMA model estimates are by far the most volatile. This corresponds to the O-EWMA conditional variance estimates being the most responsive to the observed share returns. These O-EWMA conditional correlation estimates do not appear to be reasonable as it is very unlikely that the conditional correlations would vary to that extent.

Another observation which can be made from Figure 8.2 is that the correlations in all three models appear to be slightly higher during the sub prime financial crisis. This is in line with what one expects during such a crisis for two reasons. The first is that during a financial crisis the correlations in stock markets tend to increase. Secondly the markets were fearful of the banking sector as a whole so during this time period investors tended to view the sector as a whole rather than as individual companies which also resulted in increased correlation. Therefore the models all appear to have captured this.

The results of the exchange rate data set are now considered. In the case of the exchange rate data the first plot is of the model estimates of the conditional variance of the R/£ exchange rate in Figure 8.3. Note that the range of the y-axis for the three plots of the conditional variances are all the same to facilitate comparison. These conditional variances are daily estimates and have not been annualised. The log returns of the R/£ exchange rate are also included in Figure 8.3 to get an idea of the reasonableness of the estimates. In this figure it can be observed that the majority of the conditional variance estimates are in the region of 0.124×10^{-3} which is the unconditional daily sample variance of the R/£ exchange rate returns. Therefore from that perspective the estimates appear to be reasonable. Furthermore all three models appear to capture the volatility in periods where there is increased volatility. However, unlike the share data, the O-EWMA model estimates now appear to be the least responsive out of the three to the observed returns and the O-GARCH seems to be the most responsive.

The second set of plots for the exchange rate data are of the model estimates of the conditional correlations of the R/£ and R/€ exchange rate returns and are shown in Figure 8.4. In addition a plot of the cross product of the R/£

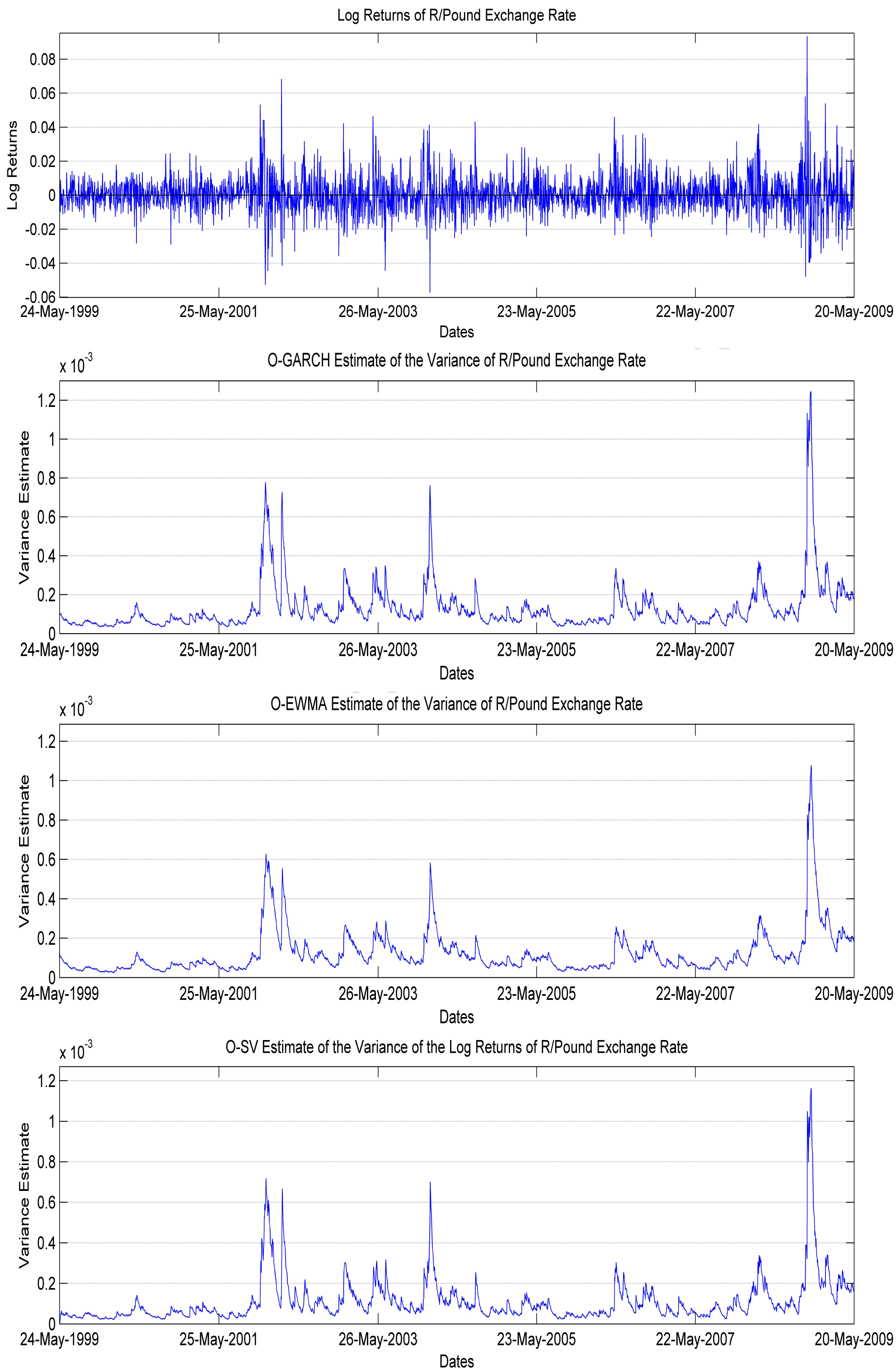


Figure 8.3: Exchange Rates: Log Returns of the R/£ and the Variance Estimates of these Returns for the O-GARCH, O-EWMA and O-SV models

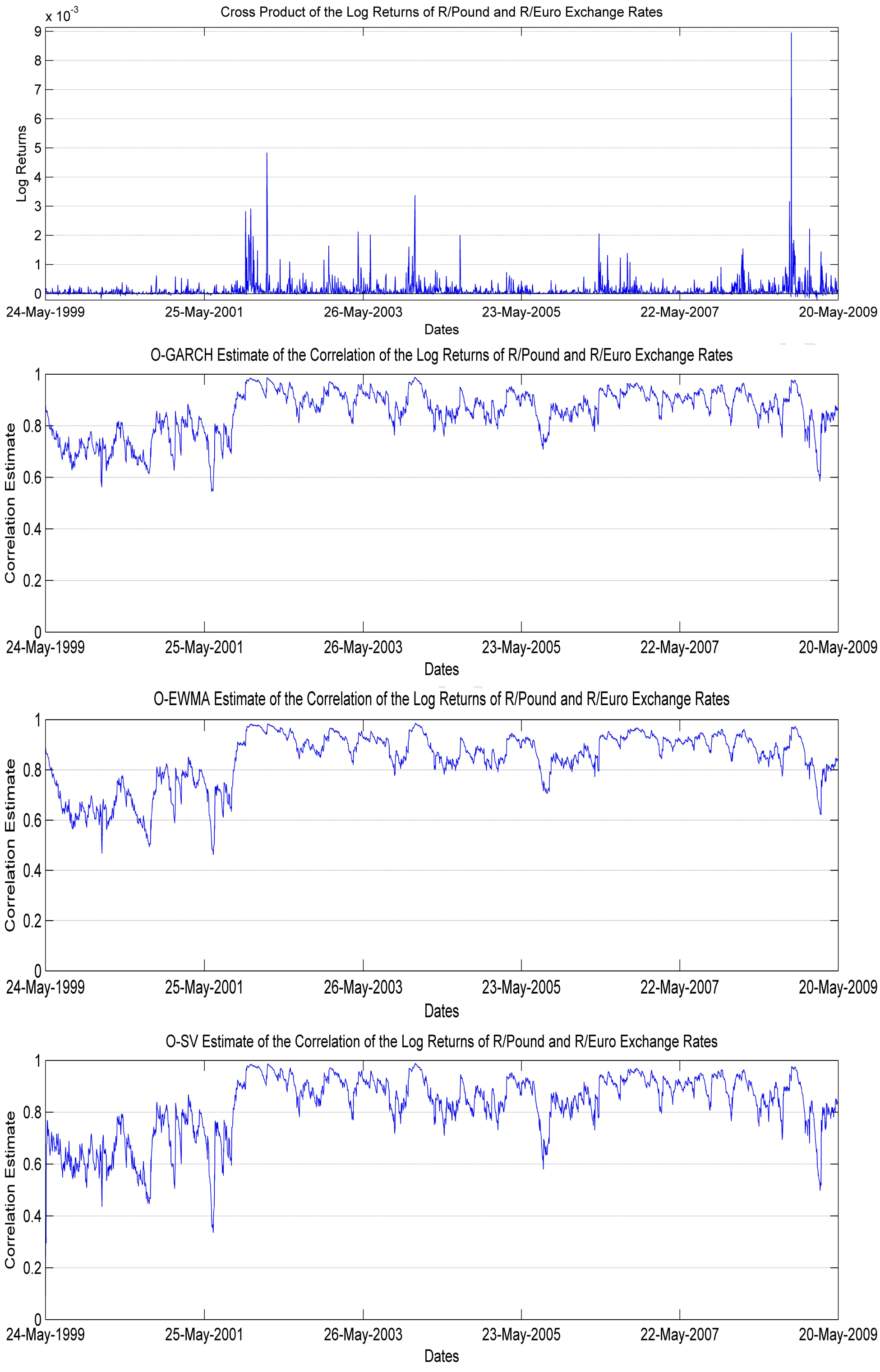


Figure 8.4: Exchange Rates: Cross Product and Correlation Estimates of the Log Returns of R/£ and R/€ Exchange Rates for the O-GARCH, O-EWMA and O-SV models

and R/€ exchange rate returns are included in Figure 8.4. In this figure the conditional correlation estimates of the three models are mostly in the region of 0.89 which is the sample correlation of the R/£ and R/€ exchange rate returns over the period (see Table 5.5). Therefore the estimates seem to be reasonable in that respect. In this case the conditional correlation estimates of the O-SV model appear to be the most responsive to the observed market returns.

Another observation which can be made is that the conditional correlation of the period from May 1999 to about August 2001 was generally lower than the rest of the period of interest. This may be due to the fact that the Euro was only introduced at the start of 1999 as a virtual currency and notes and coins only became available at the start of 2002. Therefore it was probably not as widely used initially and there was probably uncertainty about the currency when it was first launched. Both of these factors could result in a slightly lower correlation initially.

Additionally from all four Figures, 8.1 and 8.2, 8.3 and 8.4, it can be observed that the O-SV conditional covariance estimates at the start of the time period are not suitable. This is due to the choice of starting values a_{i0} and b_{i0}. Thus it may in fact be more appropriate to choose the values of a_{i0} and b_{i0} such that the conditional covariance estimate at time 1 is equal to the unconditional sample estimate. However the figures also demonstrate that even though the initial estimates are not suitable they only remain so for a very short period of time. This reinforces the point made in Chapter 7 that the choices of a_{i0} and b_{i0} do not have a big impact on the model.

8.3 Statistical Analysis of the Results

Now that an overview of the results has been discussed the model results are examined in more detail. The models in this thesis are probably best evaluated in an objective manner through the use of statistics because there are a large number of estimates output by each model.

There are three statistics which are used to evaluate the conditional covariance estimates produced by the models. These are:

1. The Box-Pierce statistic which tests for goodness of fit of each of the i^{th} and j^{th} conditional covariance estimates individually. In other words, the statistic tests the fit of the conditional covariance of the i^{th} and j^{th}

series only so a statistic is required for each combination of i and j to test the entire model, that is for $i, j = 1, 2, \cdots N$.

2. The multivariate portmanteau statistic which tests for goodness of fit of the conditional covariance model as a whole.

3. The adaptive mean absolute deviations or AMAD statistic developed by Fan et al. (2008) which tests the accuracy of the conditional covariance forecasts for the model as a whole. The statistic is the sum of the absolute errors between the conditional covariance model estimates and a proxy for the actual conditional covariances.

8.3.1 Theory

The three statistics all assess the model fit in some way. The concept behind each of these three statistics along with the formula appropriate to the orthogonal factor model output are discussed below.

8.3.1.1 Box-Pierce Statistic

The Box-Pierce statistic tests for goodness of fit of each of the i^{th} and j^{th} conditional covariance estimates individually. The particular Box-Pierce test statistic, described in this section, is used because Tse and Tsui (1999) suggest that it gives the best results when compared to the other methods tested in their paper.

Recall from equations (2.1) and (2.2) that the conditional means of the ARMA or VAR residuals $\boldsymbol{e_t}$ are zero and that the conditional covariance estimate of the residuals is $\boldsymbol{\hat{S}_t}$. Keeping to the previous notation, let $\hat{S}_{ij,t}$ be the element in the i^{th} row and j^{th} column of the matrix $\boldsymbol{\hat{S}_t}$. Therefore the sample residuals $\boldsymbol{\hat{e}_t}$ can be standardised using $\hat{S}_{ii,t}$ and $\hat{S}_{jj,t}$ to have a conditional variance of one at each point in time. Hence let $\boldsymbol{w_t} = (w_{1t}, w_{2t}, \cdots, w_{Nt})$ be an $N \times 1$ vector of the standardised residuals at time t such that w_{it} is the standardised residual of the i^{th} series so that

$$\hat{w}_{it} = \frac{\hat{e}_{it}}{(\hat{S}_{ii,t})^{\frac{1}{2}}}.$$

Consequently if the covariance matrix $\boldsymbol{\hat{S}_t}$ has correctly captured the conditional covariance of $\boldsymbol{\hat{e}_t}$ then the conditional variance of w_{it} should be 1 and the conditional covariance of w_{it} and w_{jt} should be $\frac{\hat{S}_{ij,t}}{\sqrt{\hat{S}_{ii,t}\,\hat{S}_{jj,t}}}$. Hence define

$c_{ij,t}$ as

$$c_{ij,t} = \begin{cases} w_{it}^2 - 1 & \text{if } i = j \\ w_{it}w_{jt} - \frac{\hat{S}_{ij,t}}{\sqrt{\hat{S}_{ii,t}\hat{S}_{jj,t}}} & \text{if } i \neq j. \end{cases}$$

Therefore if the model estimates $\hat{\boldsymbol{S}}_{\boldsymbol{t}}$ are close to their true value $\boldsymbol{S}_{\boldsymbol{t}}$ then the conditional mean of $c_{ij,t}$ tends to zero as the number of observations increases (Tse and Tsui, 1999). This is due to the fact that if the conditional covariance estimates $\hat{\boldsymbol{S}}_{\boldsymbol{t}}$ correctly capture the conditional covariances then the conditional mean of w_{it}^2 is 1 and the conditional mean of $w_{it}w_{jt}$ and $\frac{\hat{S}_{ij,t}}{\sqrt{\hat{S}_{ii,t}\hat{S}_{jj,t}}}$ are equal. In addition if the model is correct then asymptotically $c_{ij,t}$ should be serially uncorrelated (Tse and Tsui, 1999) for $t = 1, 2, \cdots T$.

Hence one way of testing whether the conditional covariances of the VAR or ARMA residuals are adequately modelled is to test for the presence of autocorrelation in $c_{ij,t}$. Therefore let $\varphi_{ij,k}$ be the sample autocorrelation of $c_{ij,t}$ at lag k and let $Q(i, j; k)$ be the associated Box-Pierce statistic at lag k such that

$$Q(i, j; k) = T \sum_{l=1}^{l=k} \varphi_{ij,l}^2.$$

The statistic $Q(i, j; k)$ is commonly assumed to follow an asymptotic χ_k^2 distribution, as suggested by empirical studies, although there is no theory which supports this (Tse and Tsui, 1999). Therefore this distribution will be used as a guide but not as a hard and fast rule. Despite this it is obvious that the larger the sample autocorrelations are the larger $Q(i, j; k)$ is. Therefore the larger $Q(i, j; k)$ is, the worse the fit of the covariance model.

8.3.1.2 Multivariate Portmanteau Statistic

Following the lead of Fan et al. (2008) the multivariate portmanteau statistic of Reinsel (1997) is adapted to make it suitable to test the models in this thesis. In Reinsel (1997) the statistic tests the residuals of a VAR or ARMA model to determine whether the model fit. However, in this thesis the covariances are being modelled and not the conditional mean and this statistic needs to be adapted. Fan et al. (2008) adapted the multivariate portmanteau statistic of Reinsel (1997) to make it suitable to apply to the CUC models.

Let the vector of covariance adjusted residuals be $\hat{\boldsymbol{\xi}}_{\boldsymbol{t}} = \hat{\boldsymbol{S}}_{\boldsymbol{t}}^{-\frac{1}{2}} \hat{\boldsymbol{e}}_{\boldsymbol{t}}$ where $\hat{\boldsymbol{S}}_{\boldsymbol{t}}$ is the conditional covariance estimate of the residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$. Hence the conditional covariance of this vector $\hat{\boldsymbol{\xi}}_{\boldsymbol{t}}$ should be $\boldsymbol{I}$ if the estimate $\hat{\boldsymbol{S}}_{\boldsymbol{t}}$ is the actual conditional covariance matrix at time t. Since the conditional mean of $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ is zero, $\hat{\boldsymbol{\xi}}_{\boldsymbol{t}}\hat{\boldsymbol{\xi}}_{\boldsymbol{t}}^T$ can be used as an estimate of the conditional covariance of $\hat{\boldsymbol{\xi}}_{\boldsymbol{t}}$. Therefore applying the statistic of Reinsel (1997) to $\hat{\boldsymbol{\xi}}_{\boldsymbol{t}}\hat{\boldsymbol{\xi}}_{\boldsymbol{t}}^T$ to test for goodness of fit of the conditional covariance models is analogous to applying the statistic to test for goodness of fit of a VAR or ARMA conditional mean model. However $\hat{\boldsymbol{\xi}}_{\boldsymbol{t}}\hat{\boldsymbol{\xi}}_{\boldsymbol{t}}^T$ is a matrix and the multivariate portmanteau statistic in Reinsel (1997) tests a vector of residuals and not a matrix. Thus to overcome this problem the matrix is vectorised to give $\boldsymbol{\Upsilon}_{\boldsymbol{t}} = vech(\hat{\boldsymbol{\xi}}_{\boldsymbol{t}}\hat{\boldsymbol{\xi}}_{\boldsymbol{t}}^T)$. This vector $\boldsymbol{\Upsilon}_{\boldsymbol{t}}$ is now treated as the residuals of a VAR or ARMA model in the multivariate portmanteau statistic of Reinsel (1997).

Let

$$\boldsymbol{C}_{\Upsilon}(l) = \frac{1}{T}\sum_{t=1}^{T-l}(\boldsymbol{\Upsilon}_{\boldsymbol{t}})(\boldsymbol{\Upsilon}_{t+l})^T.$$

This gives an estimate of the covariance matrix of $\boldsymbol{\Upsilon}_{\boldsymbol{t}}$ when $l = 0$. Therefore the overall goodness of fit statistic of Reinsel (1997) testing up to lag k is given by

$$Q_k = T^2\sum_{l=1}^{k}(T-l)^{-1}tr\{\boldsymbol{C}_{\Upsilon}(l)\boldsymbol{C}_{\Upsilon}(0)^{-1}\boldsymbol{C}_{\Upsilon}(-l)\boldsymbol{C}_{\Upsilon}(0)^{-1}\}$$

where T is the sample size. Consequently the i^{th} and j^{th} element of the matrix $\boldsymbol{C}_{\Upsilon}(l)\boldsymbol{C}_{\Upsilon}(0)^{-1}$ is analogous to the residual correlation of the i^{th} and j^{th} series. Therefore Q_k is comparable to the sum of the squared residual correlations up to lag k.

The distribution of the statistic Q_k is unknown but it is intuitive that the larger the values of Q_k are the poorer the model fit is (Fan et al., 2008). This is because the residual correlations should be small if the conditional covariance has been well modelled.

8.3.1.3 AMAD Statistic

The AMAD (adaptive mean absolute deviations) statistic is a measure of the predictive power of a model and is discussed in Fan et al. (2008). The statistic is calculated as the sum over time of the absolute errors of the forecasted conditional covariances of the ARMA or VAR residuals $\boldsymbol{\hat{e}_t}$. Therefore if the c-step ahead forecast is being evaluated then the statistic is a function of c and will be denoted as AMAD(c).

Hence let $\hat{S}_{ij,\,t+c|t}$ be the c-step ahead forecast of the conditional sample covariance of the i^{th} and j^{th} series of the estimated residuals given information up to time and including time t. However the actual conditional covariance matrix required to calculate the statistic cannot be observed so a proxy is necessary. AMAD proxies the conditional covariance of the i^{th} and j^{th} series of residuals at time t as

$$\frac{1}{1+2v}\sum_{l=-v}^{l=v} \hat{e}_{i,t+l}\,\hat{e}_{j,t+l}$$

where $\hat{e}_{it}$ is the i^{th} series of the estimated residual $\boldsymbol{\hat{e}_t}$ at time t. This proxy was discussed in more detail in Section 8.1.

Therefore if the accuracy of the c-step ahead forecast is being tested then the AMAD(c) statistic is given by

$$AMAD(c) \;=\; \frac{1}{N^2 M}\sum_{t}\sum_{i,j=1}^{N}\left|\hat{S}_{ij,\,t+c|t} - \frac{1}{2v+1}\sum_{l=-v}^{l=v}\hat{e}_{i,t+l}\,\hat{e}_{j,t+l}\right|$$

where M is the number of time points t which over which the sum is taken. Therefore a rolling window is used such that each time the window is rolled forward the model parameters are recalculated and a forecast is calculated based on those model parameters. The AMAD statistics calculated in this thesis are rolled forward 250 times, in other words $M = 250$. This implies that the forecasts are tested over the period of approximately one year.

The AMAD statistic is also a function of the period of time of the window which is rolled forward, although this is not obvious from the formula. In this thesis 500 and 1000 data points are used in this statistic to calculate the model parameters. This is represented by the symbol T_A in the tables containing the AMAD statistics as this is the number of sample points used to calculate the statistic. Therefore if 1000 points are used, that is $T_A = 1000$, then the time points t to $t + 1000$ are used to estimate the model parameters and this is used to calculate a c-step ahead forecast. Then time points $t + 1$ to $t + 1001$ are used to recalculate the model parameters and from this a new c-step ahead forecast is calculated and so forth.

Although there is no known distribution for the AMAD statistic it is obvious that the larger the statistic, the worse the forecasts are (Fan et al., 2008).

8.3.2 Applications and Comparisons of the Statistics

Each of the three statistics, discussed in the previous section, have different advantages and disadvantages and therefore the three statistics together can help gain greater insight into the results. A brief discussion of these advantages and disadvantages are presented here.

The multivariate portmanteau and AMAD statistics both have the advantage that an entire model's results are summarised into one number. This facilitates easy comparison and evaluation of the models. On the other hand the disadvantage of these statistics is that valuable information may be lost as the statistic is only one number. However, the AMAD statistic can do more than just evaluate a model. It can also be used to help "tweak" the model to obtain more accurate forecasts. For example it can help to find the best time period to use for estimating the model parameters and whether the forecasts are better for estimating an individual point or a region around a point.

On the other hand the advantages and disadvantages of the Box-Pierce statistic are the exact opposite to the advantages and disadvantages of the multivariate portmanteau and AMAD statistics. Since the Box-Pierce statistic does not summarise the results of the entire model into a single number it provides additional information that can help to determine where the weaknesses are in the model. For example it gives an indication of the extent of the negative impact on the results due to the orthogonal matrix $\boldsymbol{A}$ not removing the conditional cross correlations, as assumed in the model. This is achieved by calculating the Box-Pierce statistic for the principal component scores to determine whether the Box-Pierce statistics are on average greater when $i \neq j$ than when $i = j$. Although the theory described is for the Box-Pierce statistic of the ARMA or VAR residuals it can easily be adapted so that it can be calculated for the principal component scores instead.

Therefore these three statistics together can help to gain insight into the models, and into the best form to use for each model, which of the three models are best overall and to find some of the model weaknesses or strengths. Although no definite conclusions can be made from these statistics they may be able to give an indication of what may be true in general. Thus it is possible to apply the methods in this thesis to a range of other data sets to further investigate whether the findings are true more generally.

8.3.3 Results of the Statistics

8.3.3.1 Box-Pierce Statistic Results

The Box-Pierce statistic $Q(i,j;k)$ requires a lag k to be selected. It is preferable to choose k to be at least 5 as this will identify any daily as well as weekly effects not captured correctly by a model. Therefore $k = 5$ is chosen as this ensures that a sufficient number of lags are tested and that the test has sufficient power.

The statistics can be found in Tables 8.1 to 8.16. These tables contain Q-statistics which are fitted to the ARMA or VAR sample residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ and Q-statistics which are fitted to the principal component scores $\boldsymbol{x_t}$. Tables 8.1 to 8.6 contain the Q-statistics for the O-GARCH models, Tables 8.7 to 8.10 contain the Q-statistics for the O-EWMA models and Tables 8.11 to 8.16 contain the Q-statistics for the O-SV models. The O-EWMA Tables 8.7 to 8.10 do not contain $\boldsymbol{W}$ as the two variations of $\boldsymbol{W}$ give identical results for this model.

Note that under the assumption that $Q(i,j;k)$ does in fact follow a χ^2_k distribution the cells highlighted in dark blue indicate significance at the 1% level, cells in medium blue indicate significance at the 5% level and lastly cells in light blue at the 10% significance level. However to interpret this as significance is not quite correct as the distribution is not necessarily accurate. The cells are merely highlighted to give an indication of which statistics are large and to allow an overview of the results from a quick look at the tables.

There are a number of comparisons which can be made using the Q-statistics. Firstly for each of the O-GARCH, O-EWMA and O-SV models the variations, that is the choice of conditional mean model, $\boldsymbol{B}$ and $\boldsymbol{W}$, are compared by keeping two of the three fixed and looking for any "trends" in the Q-statistics by changing the third. Therefore if the Q-statistics of the one choice is usually smaller than the other then that choice may well be preferable. Secondly the three models O-GARCH, O-EWMA and O-SV can be compared by determining whether the Q-statistics of one of the models are generally smaller than the others, keeping all else constant. This is essentially more of a qualitative comparison than a quantitative comparison in the sense that it is an overall look at the results. Thirdly the Q-statistics

Table 8.1: Shares: $Q(i, j; 5)$ of the Sample Residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ of the O-GARCH Model, modelling $\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$ as ARMA

$\boldsymbol{B}$ $\boldsymbol{W}$	$\boldsymbol{I}$ $\boldsymbol{I}$	$\boldsymbol{I}$ $(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ $\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ $(\boldsymbol{\Lambda})^{-\frac{1}{2}}$
$i = 1,\ j = 1$	26.7941	26.7407	25.2306	25.2302
$i = 2,\ j = 2$	18.0467	18.0297	17.6435	17.6437
$i = 3,\ j = 3$	32.5740	32.4318	29.6006	29.6013
$i = 4,\ j = 4$	33.0373	32.9420	31.6790	31.6789
$i = 5,\ j = 5$	7.7500	7.7416	22.2565	22.2564
$i = 6,\ j = 6$	35.3420	35.3304	48.2325	48.2329
$i = 7,\ j = 7$	22.2398	22.1825	37.9052	37.9075
$i = 1,\ j = 2$	3.3824	3.3792	3.4677	3.4678
$i = 1,\ j = 3$	5.0464	5.0447	5.6154	5.6154
$i = 1,\ j = 4$	2.1995	2.1824	1.7132	1.7132
$i = 1,\ j = 5$	25.5282	25.5298	26.7417	26.7415
$i = 1,\ j = 6$	8.1227	8.1149	8.6185	8.6185
$i = 1,\ j = 7$	3.1471	3.1425	2.9950	2.9951
$i = 2,\ j = 3$	12.7151	12.7064	12.2024	12.2025
$i = 2,\ j = 4$	3.3659	3.3598	3.1748	3.1748
$i = 2,\ j = 5$	11.1406	11.1394	14.1422	14.1421
$i = 2,\ j = 6$	9.2666	9.2684	11.9040	11.9036
$i = 2,\ j = 7$	1.5396	1.5358	1.3950	1.3950
$i = 3,\ j = 4$	3.5567	3.5532	3.0898	3.0898
$i = 3,\ j = 5$	8.3354	8.3398	8.8071	8.8070
$i = 3,\ j = 6$	9.3221	9.3292	9.2060	9.2059
$i = 3,\ j = 7$	3.6135	3.6111	2.6818	2.6817
$i = 4,\ j = 5$	11.7443	11.7330	11.9017	11.9015
$i = 4,\ j = 6$	15.2019	15.2098	16.7560	16.7559
$i = 4,\ j = 7$	7.2187	7.2208	7.5296	7.5296
$i = 5,\ j = 6$	6.5434	6.5441	19.0204	19.0206
$i = 5,\ j = 7$	10.2009	10.2180	10.6735	10.6732
$i = 6,\ j = 7$	14.6950	14.6890	14.9481	14.9480

Table 8.2: Shares: $Q(i, j; 5)$ of the Sample Residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ of the O-GARCH Model, modelling $\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$ as VAR

$\boldsymbol{B}$ $\boldsymbol{W}$	$\boldsymbol{I}$ $\boldsymbol{I}$	$\boldsymbol{I}$ $(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ $\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ $(\boldsymbol{\Lambda})^{-\frac{1}{2}}$
$i = 1,\ j = 1$	29.9485	29.9428	29.2197	29.2203
$i = 2,\ j = 2$	14.9342	14.8853	15.1335	15.1342
$i = 3,\ j = 3$	31.8527	31.6823	28.9771	28.9786
$i = 4,\ j = 4$	35.2449	35.1940	34.2642	34.2642
$i = 5,\ j = 5$	6.2956	6.2761	24.6779	24.6782
$i = 6,\ j = 6$	35.7494	35.7958	51.4654	51.4655
$i = 7,\ j = 7$	21.0424	20.9980	32.1384	32.1374
$i = 1,\ j = 2$	2.2656	2.2655	2.5829	2.5828
$i = 1,\ j = 3$	5.4461	5.4439	6.1223	6.1222
$i = 1,\ j = 4$	1.9448	1.9322	1.8160	1.8160
$i = 1,\ j = 5$	22.3171	22.3043	25.5255	25.5256
$i = 1,\ j = 6$	10.8288	10.8297	10.4833	10.4833
$i = 1,\ j = 7$	3.5336	3.5345	3.3438	3.3438
$i = 2,\ j = 3$	15.4335	15.4155	15.5325	15.5325
$i = 2,\ j = 4$	3.3487	3.3401	3.5012	3.5013
$i = 2,\ j = 5$	9.4956	9.4919	13.1385	13.1386
$i = 2,\ j = 6$	11.3318	11.3486	17.1269	17.1268
$i = 2,\ j = 7$	5.1634	5.1587	5.0901	5.0905
$i = 3,\ j = 4$	4.5820	4.5864	4.2137	4.2138
$i = 3,\ j = 5$	8.7486	8.7530	10.5166	10.5167
$i = 3,\ j = 6$	6.0887	6.0961	5.7326	5.7328
$i = 3,\ j = 7$	4.6985	4.6970	3.8966	3.8969
$i = 4,\ j = 5$	11.9536	11.9271	12.5195	12.5196
$i = 4,\ j = 6$	15.1220	15.1372	17.9681	17.9684
$i = 4,\ j = 7$	8.1025	8.1036	8.1516	8.1519
$i = 5,\ j = 6$	13.9002	13.9013	20.6882	20.6882
$i = 5,\ j = 7$	10.4834	10.4945	12.7748	12.7752
$i = 6,\ j = 7$	14.1957	14.1970	13.9402	13.9402

Table 8.3: Shares: $Q(i, j; 5)$ of the Principal Component Scores $\boldsymbol{x_t}$ of the O-GARCH Model, modelling $\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$ as ARMA

$\boldsymbol{B}$ $\boldsymbol{W}$	$\boldsymbol{I}$ $\boldsymbol{I}$	$\boldsymbol{I}$ $(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ $\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ $(\boldsymbol{\Lambda})^{-\frac{1}{2}}$
$i = 1,\ j = 1$	12.4809	12.4682	13.1509	13.1510
$i = 2,\ j = 2$	8.3980	8.3909	6.6358	6.6358
$i = 3,\ j = 3$	16.3073	16.2733	15.1227	15.1205
$i = 4,\ j = 4$	8.8364	8.7911	12.4832	12.4840
$i = 5,\ j = 5$	11.8394	11.7651	11.5691	11.5690
$i = 6,\ j = 6$	11.3712	11.0724	11.6046	11.6034
$i = 7,\ j = 7$	24.0987	23.8576	24.0465	24.0490
$i = 1,\ j = 2$	24.3425	24.3403	36.4374	36.4374
$i = 1,\ j = 3$	26.7589	26.7670	22.8708	22.8708
$i = 1,\ j = 4$	6.7757	6.7807	8.5225	8.5224
$i = 1,\ j = 5$	9.6394	9.6094	8.7879	8.7880
$i = 1,\ j = 6$	12.5335	12.5068	13.6726	13.6725
$i = 1,\ j = 7$	2.2212	2.2182	1.7644	1.7644
$i = 2,\ j = 3$	1.4882	1.4878	21.4370	21.4373
$i = 2,\ j = 4$	4.2682	4.2712	5.7264	5.7263
$i = 2,\ j = 5$	0.6474	0.6480	3.0880	3.0880
$i = 2,\ j = 6$	9.5476	9.5313	9.5860	9.5859
$i = 2,\ j = 7$	2.8597	2.8682	3.8874	3.8874
$i = 3,\ j = 4$	4.1354	4.1337	12.5229	12.5231
$i = 3,\ j = 5$	0.8740	0.8589	1.5656	1.5655
$i = 3,\ j = 6$	6.0431	6.0010	7.6412	7.6409
$i = 3,\ j = 7$	8.1493	8.1573	5.6100	5.6098
$i = 4,\ j = 5$	7.2596	7.2434	8.9694	8.9694
$i = 4,\ j = 6$	4.6739	4.6838	4.2346	4.2346
$i = 4,\ j = 7$	3.2741	3.2893	2.5567	2.5567
$i = 5,\ j = 6$	61.4466	61.4708	61.1636	61.1633
$i = 5,\ j = 7$	3.5462	3.5742	3.7003	3.7002
$i = 6,\ j = 7$	40.4623	40.4558	38.9584	38.9584

Table 8.4: Shares: $Q(i,j;5)$ of the Principal Component Scores $\boldsymbol{x_t}$ of the O-GARCH Model, modelling $\boldsymbol{\hat{\mu}_t}$ as VAR

$\boldsymbol{B}$ / $\boldsymbol{W}$	$\boldsymbol{I}$ / $\boldsymbol{I}$	$\boldsymbol{I}$ / $(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ / $\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ / $(\boldsymbol{\Lambda})^{-\frac{1}{2}}$
$i=1,\ j=1$	13.0289	13.0173	13.4912	13.4915
$i=2,\ j=2$	7.4724	7.4553	8.3487	8.3488
$i=3,\ j=3$	14.4548	14.4406	46.1796	46.1798
$i=4,\ j=4$	8.5478	8.5112	7.9853	7.9848
$i=5,\ j=5$	11.8527	11.7942	11.6623	11.6614
$i=6,\ j=6$	5.0855	4.9275	5.1347	5.1364
$i=7,\ j=7$	17.9185	17.5739	17.9115	17.9119
$i=1,\ j=2$	20.6841	20.6789	31.2690	31.2692
$i=1,\ j=3$	22.2693	22.2773	21.7018	21.7018
$i=1,\ j=4$	5.1910	5.1941	7.0658	7.0659
$i=1,\ j=5$	9.7579	9.7393	8.6917	8.6918
$i=1,\ j=6$	14.0695	14.0447	15.3930	15.3935
$i=1,\ j=7$	3.3449	3.3360	2.9013	2.9012
$i=2,\ j=3$	4.3069	4.3043	20.9635	20.9635
$i=2,\ j=4$	4.4234	4.4270	9.1515	9.1515
$i=2,\ j=5$	0.8793	0.8959	4.3393	4.3393
$i=2,\ j=6$	10.5750	10.5542	9.5897	9.5899
$i=2,\ j=7$	5.0064	5.0195	5.7448	5.7448
$i=3,\ j=4$	4.0667	4.0635	9.2849	9.2848
$i=3,\ j=5$	0.5395	0.5321	0.4786	0.4785
$i=3,\ j=6$	6.4652	6.4404	6.8803	6.8808
$i=3,\ j=7$	6.3902	6.4000	3.6368	3.6368
$i=4,\ j=5$	5.8921	5.8905	9.2946	9.2943
$i=4,\ j=6$	1.6497	1.6612	2.1937	2.1937
$i=4,\ j=7$	3.5673	3.5806	4.2029	4.2029
$i=5,\ j=6$	49.6530	49.7385	50.8658	50.8662
$i=5,\ j=7$	4.0781	4.1008	4.2063	4.2063
$i=6,\ j=7$	29.0019	29.0080	29.0334	29.0335

Table 8.5: Exchange Rates: $Q(i, j; 5)$ of the Sample Residuals $\hat{\boldsymbol{e}}_t$ of the O-GARCH Model

$\hat{\boldsymbol{\mu}}_t$	ARMA	ARMA	ARMA	ARMA	VAR	VAR	VAR	VAR
$\boldsymbol{B}$	$\boldsymbol{I}$	$\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	$\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$
$\boldsymbol{W}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$
$i = 1,\ j = 1$	10.6613	10.7991	12.0290	12.0287	8.8976	9.0392	9.4626	9.7817
$i = 2,\ j = 2$	2.4757	2.4374	2.3290	2.3289	2.6769	2.6228	1.8681	2.3576
$i = 3,\ j = 3$	21.8671	21.9915	26.6392	26.6389	20.3315	20.4173	24.3448	24.3381
$i = 4,\ j = 4$	3.1951	3.1880	2.1033	2.1033	3.6487	3.6086	2.2644	2.2892
$i = 5,\ j = 5$	43.0158	42.8181	57.6317	57.6303	38.8825	38.9101	50.7526	52.1652
$i = 1,\ j = 2$	1.4091	1.3753	1.6367	1.6366	1.6262	1.5652	1.7121	1.8187
$i = 1,\ j = 3$	2.4485	2.3483	2.4873	2.4871	1.7725	1.6820	1.8353	1.8008
$i = 1,\ j = 4$	1.5072	1.5122	1.1301	1.1302	1.8589	1.8508	1.0149	1.0960
$i = 1,\ j = 5$	4.5417	4.5473	7.2486	7.2482	4.0225	4.0173	5.9991	5.9990
$i = 2,\ j = 3$	1.9969	2.0036	1.8052	1.8050	1.2280	1.2373	1.0511	1.0586
$i = 2,\ j = 4$	1.5483	1.5484	1.5058	1.5058	2.0443	2.0357	1.1823	1.4035
$i = 2,\ j = 5$	3.5677	3.5262	5.5079	5.5076	4.0562	4.0040	5.8622	5.8545
$i = 3,\ j = 4$	2.3787	2.3855	2.8793	2.8795	1.3635	1.3707	1.8095	1.8555
$i = 3,\ j = 5$	9.8694	9.7771	16.1711	16.1703	9.0362	8.9736	14.9239	15.1464
$i = 4,\ j = 5$	2.6882	2.7326	3.3696	3.3698	4.9111	4.9560	5.6254	5.6916

Table 8.6: Exchange Rates: $Q(i,j;5)$ of the Principal Component Scores $\boldsymbol{x_t}$ of the O-GARCH Model

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$	ARMA	ARMA	ARMA	ARMA	VAR	VAR	VAR	VAR
$\boldsymbol{B}$	$\boldsymbol{I}$	$\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	$\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$
$\boldsymbol{W}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$
$i=1,\ j=1$	0.4136	0.4131	0.3107	0.3107	0.5263	0.5270	0.4271	0.4271
$i=2,\ j=2$	4.2829	4.1410	3.8871	3.8868	6.2935	6.3260	4.5872	4.5888
$i=3,\ j=3$	5.2156	5.3579	3.9357	3.9356	5.2152	5.3981	17.9724	2.9287
$i=4,\ j=4$	2.6544	2.6538	2.0513	2.0512	5.3134	5.3107	4.2981	4.2981
$i=5,\ j=5$	5.2261	5.6404	4.0858	4.0857	4.8233	4.7764	4.2848	4.2850
$i=1,\ j=2$	107.5093	107.4680	77.3659	77.3655	119.2255	119.4460	86.0751	86.0758
$i=1,\ j=3$	36.9663	37.3908	31.7058	31.7052	27.7334	27.9436	22.9269	27.8045
$i=1,\ j=4$	69.0169	68.9945	60.6243	60.6246	86.0405	86.0121	68.3521	68.3508
$i=1,\ j=5$	8.8703	9.0053	9.9692	9.9693	13.2612	13.1780	18.0850	18.0843
$i=2,\ j=3$	37.4216	37.5575	31.6026	31.6024	42.4548	42.6370	34.6151	37.4367
$i=2,\ j=4$	7.0301	7.0139	17.1831	17.1832	8.1707	8.1576	13.8718	13.8719
$i=2,\ j=5$	17.0323	15.2407	18.4241	18.4241	13.8469	11.7179	14.7766	14.7766
$i=3,\ j=4$	14.0471	14.1099	16.2501	16.2500	10.1840	10.2328	10.6344	10.4127
$i=3,\ j=5$	40.3028	37.8307	67.0291	67.0289	34.3748	32.0608	58.6255	60.4315
$i=4,\ j=5$	52.5077	53.2878	44.8495	44.8498	55.4521	55.3354	55.8906	55.8906

Table 8.7: Shares: $Q(i, j; 5)$ of the Sample Residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ of the O-EWMA Model

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$	ARMA	ARMA	VAR	VAR
$\boldsymbol{B}$	$\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$
$i = 1,\ j = 1$	1.3763	1.1088	1.8310	1.8660
$i = 2,\ j = 2$	10.4995	7.5797	8.9027	6.8969
$i = 3,\ j = 3$	28.1494	25.7834	18.7293	15.2633
$i = 4,\ j = 4$	7.1214	9.6129	9.7590	15.0152
$i = 5,\ j = 5$	11.1400	28.6516	10.7455	29.8216
$i = 6,\ j = 6$	6.8860	11.6933	5.5398	9.7845
$i = 7,\ j = 7$	12.4535	16.9598	13.5737	20.1432
$i = 1,\ j = 2$	22.6168	22.3357	21.4351	19.5908
$i = 1,\ j = 3$	29.9990	31.3608	28.4916	26.2878
$i = 1,\ j = 4$	29.4683	27.9762	23.2762	18.2268
$i = 1,\ j = 5$	15.5732	19.3573	15.2513	19.3711
$i = 1,\ j = 6$	17.7979	12.5488	13.3271	6.8295
$i = 1,\ j = 7$	14.8646	14.4713	10.6534	7.7315
$i = 2,\ j = 3$	21.9316	20.4412	21.9337	19.5621
$i = 2,\ j = 4$	21.3323	18.8012	18.5695	16.1659
$i = 2,\ j = 5$	4.5575	8.4013	2.7999	6.2779
$i = 2,\ j = 6$	19.7683	12.8159	14.3025	8.3522
$i = 2,\ j = 7$	4.5384	4.9390	3.2191	2.4246
$i = 3,\ j = 4$	38.7128	34.9964	33.0105	25.5536
$i = 3,\ j = 5$	6.7894	8.8084	6.2533	8.5284
$i = 3,\ j = 6$	9.1140	5.3503	5.8717	2.6630
$i = 3,\ j = 7$	7.2584	6.5755	4.5115	3.0545
$i = 4,\ j = 5$	8.2163	9.9476	10.4468	11.7495
$i = 4,\ j = 6$	16.1518	11.7558	10.4910	8.2249
$i = 4,\ j = 7$	8.4924	7.8653	3.6298	2.8769
$i = 5,\ j = 6$	5.5071	13.2833	11.3057	19.0236
$i = 5,\ j = 7$	3.9211	6.0061	5.2062	8.4656
$i = 6,\ j = 7$	9.5868	8.7582	8.2865	7.8790

Table 8.8: Shares: $Q(i,j;5)$ of the Principal Component Scores $\boldsymbol{x_t}$ of the O-EWMA Model

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$ / $\boldsymbol{B}$	ARMA $\boldsymbol{I}$	ARMA $diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	VAR $\boldsymbol{I}$	VAR $diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$
$i=1,\ j=1$	11.0106	11.0209	11.0142	11.0423
$i=2,\ j=2$	10.8569	10.7674	10.9390	11.0333
$i=3,\ j=3$	10.9044	10.9345	11.0243	11.0209
$i=4,\ j=4$	10.9395	11.0387	10.9453	11.0325
$i=5,\ j=5$	10.8198	10.8248	10.5793	10.9538
$i=6,\ j=6$	10.8199	10.8719	6.5464	6.9032
$i=7,\ j=7$	11.0408	11.0234	11.0164	10.9564
$i=1,\ j=2$	1.4931	3.7214	1.4428	5.9482
$i=1,\ j=3$	10.7375	11.2599	10.7082	12.5514
$i=1,\ j=4$	4.7149	4.8851	2.2024	3.0527
$i=1,\ j=5$	2.8168	3.9051	2.6148	3.4231
$i=1,\ j=6$	5.7831	6.4897	5.9587	8.4893
$i=1,\ j=7$	0.6997	0.5904	0.8674	1.4497
$i=2,\ j=3$	1.3289	15.3826	4.5925	20.9294
$i=2,\ j=4$	4.5852	4.4197	4.3768	7.5514
$i=2,\ j=5$	1.0149	2.8067	1.2644	3.8352
$i=2,\ j=6$	8.1506	8.8926	8.9328	8.9629
$i=2,\ j=7$	0.4521	1.7104	1.0872	2.9144
$i=3,\ j=4$	3.4084	7.6607	4.3136	7.2473
$i=3,\ j=5$	2.2472	2.4505	2.0098	1.4684
$i=3,\ j=6$	9.1082	10.5942	6.6798	7.2709
$i=3,\ j=7$	4.8877	3.5566	3.7709	2.1700
$i=4,\ j=5$	6.6032	9.0947	5.2907	9.7525
$i=4,\ j=6$	4.1749	4.9013	1.0551	2.0063
$i=4,\ j=7$	2.4416	2.3792	3.8643	4.0968
$i=5,\ j=6$	44.3716	43.5689	37.5472	38.6172
$i=5,\ j=7$	0.9803	1.3448	1.4990	1.5098
$i=6,\ j=7$	14.6202	14.1037	10.7065	10.2587

Table 8.9: Exchange Rates: $Q(i, j; 5)$ of the Sample Residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ of the O-EWMA Model

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$ $\boldsymbol{B}$	ARMA $\boldsymbol{I}$	ARMA $diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	VAR $\boldsymbol{I}$	VAR $diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$
$i = 1,\ j = 1$	9.0998	9.7951	9.1788	9.6216
$i = 2,\ j = 2$	3.6433	3.2834	4.9382	4.5135
$i = 3,\ j = 3$	13.8221	15.6580	12.6084	14.2304
$i = 4,\ j = 4$	6.7411	4.9602	8.2493	6.2903
$i = 5,\ j = 5$	31.9700	39.9672	27.9224	35.6367
$i = 1,\ j = 2$	3.0403	3.0256	4.0686	3.9777
$i = 1,\ j = 3$	4.5706	4.4206	4.0389	4.0235
$i = 1,\ j = 4$	4.5506	3.1306	5.9294	4.4326
$i = 1,\ j = 5$	4.6291	6.4500	4.5904	6.3879
$i = 2,\ j = 3$	3.9920	3.6098	3.7087	3.2868
$i = 2,\ j = 4$	3.5224	2.3632	4.9815	3.5645
$i = 2,\ j = 5$	3.4959	5.0277	3.9624	5.4561
$i = 3,\ j = 4$	4.1487	3.0159	3.5698	2.4666
$i = 3,\ j = 5$	9.8874	12.6343	8.4178	11.0867
$i = 4,\ j = 5$	2.8737	2.6575	3.3878	3.4007

Table 8.10: Exchange Rates: $Q(i, j; 5)$ of the Principal Component Scores $\boldsymbol{x_t}$ of the O-EWMA Model

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$ $\boldsymbol{B}$	ARMA $\boldsymbol{I}$	ARMA $diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	VAR $\boldsymbol{I}$	VAR $diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$
$i=1,\ j=1$	1.9330	1.8681	1.8281	1.8481
$i=2,\ j=2$	9.8778	6.5920	10.7143	7.4107
$i=3,\ j=3$	6.1218	4.2226	7.5335	3.3142
$i=4,\ j=4$	2.6512	1.9432	6.6784	4.2206
$i=5,\ j=5$	5.6686	3.3132	4.0571	2.7638
$i=1,\ j=2$	68.2539	52.6307	80.6067	63.2781
$i=1,\ j=3$	36.7576	30.8110	29.2611	26.3576
$i=1,\ j=4$	61.9425	56.6722	77.9423	64.1899
$i=1,\ j=5$	8.9200	10.0090	12.7545	17.8886
$i=2,\ j=3$	30.8656	22.3473	32.8554	26.4591
$i=2,\ j=4$	8.8961	17.0135	10.6093	13.8825
$i=2,\ j=5$	8.9556	12.4449	5.1658	7.8933
$i=3,\ j=4$	13.1924	15.1909	9.2452	8.7523
$i=3,\ j=5$	25.4415	53.6018	21.5658	47.7915
$i=4,\ j=5$	44.4367	32.4424	43.0809	38.2474

Table 8.11: Shares: $Q(i, j; 5)$ of the Sample Residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ of the O-SV Model, modelling $\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$ as ARMA

$\boldsymbol{B}$	$\boldsymbol{I}$	$\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$
$\boldsymbol{W}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$
$i = 1,\ j = 1$	9.3230	5.9788	7.9812	5.2515
$i = 2,\ j = 2$	21.3025	19.8935	18.2331	17.3317
$i = 3,\ j = 3$	49.6794	45.9770	43.3288	46.1153
$i = 4,\ j = 4$	37.3219	40.3113	36.9228	38.5261
$i = 5,\ j = 5$	12.8065	0.0002	0.0017	0.0016
$i = 6,\ j = 6$	15.3092	0.4687	11.4373	9.3413
$i = 7,\ j = 7$	24.2913	24.0389	37.2820	37.2670
$i = 1,\ j = 2$	7.8390	3.6277	4.9474	2.5798
$i = 1,\ j = 3$	10.9568	13.1835	14.5839	14.2450
$i = 1,\ j = 4$	6.8798	7.6593	6.9122	7.9111
$i = 1,\ j = 5$	22.6806	0.0061	0.0868	0.0328
$i = 1,\ j = 6$	4.1245	0.6666	3.0291	1.5722
$i = 1,\ j = 7$	4.2531	2.2232	2.9324	2.6560
$i = 2,\ j = 3$	24.9599	13.8085	18.0127	11.0860
$i = 2,\ j = 4$	15.7234	13.6013	12.2004	10.8291
$i = 2,\ j = 5$	5.5342	0.0097	0.5277	0.1229
$i = 2,\ j = 6$	8.1213	2.0891	7.3166	5.0691
$i = 2,\ j = 7$	6.9174	6.8843	5.3187	5.2755
$i = 3,\ j = 4$	14.7682	15.4889	14.8462	15.4381
$i = 3,\ j = 5$	7.0093	0.0062	0.0890	0.0310
$i = 3,\ j = 6$	7.9685	1.2703	5.1483	2.3885
$i = 3,\ j = 7$	6.9492	1.4135	2.9194	1.8225
$i = 4,\ j = 5$	15.8367	0.0545	0.7012	0.3296
$i = 4,\ j = 6$	14.3888	5.4704	12.3997	9.5954
$i = 4,\ j = 7$	6.5056	5.6814	5.6298	5.5955
$i = 5,\ j = 6$	5.3263	0.0008	0.0185	0.0147
$i = 5,\ j = 7$	5.1771	0.1253	1.3681	1.1437
$i = 6,\ j = 7$	13.0584	6.6515	11.9049	11.5472

Table 8.12: Shares: $Q(i, j; 5)$ of the Sample Residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ of the O-SV Model, modelling $\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$ as VAR

$\boldsymbol{B}$	$\boldsymbol{I}$	$\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$
$\boldsymbol{W}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$
$i = 1,\ j = 1$	11.8801	8.5102	11.0686	7.9334
$i = 2,\ j = 2$	19.6790	18.7896	16.7149	16.0036
$i = 3,\ j = 3$	47.7319	40.5472	37.9940	38.2120
$i = 4,\ j = 4$	39.2273	43.0382	39.9293	41.7851
$i = 5,\ j = 5$	12.1485	0.0002	0.0021	0.0020
$i = 6,\ j = 6$	17.4236	5.8293	16.2173	14.4455
$i = 7,\ j = 7$	24.0083	23.4956	35.4081	35.3846
$i = 1,\ j = 2$	6.3451	3.6109	4.2306	2.4946
$i = 1,\ j = 3$	12.5512	12.2942	12.5352	12.9196
$i = 1,\ j = 4$	6.1679	7.5240	6.0107	7.6136
$i = 1,\ j = 5$	20.7817	0.0081	0.1402	0.0510
$i = 1,\ j = 6$	4.2016	2.5303	4.6129	2.9726
$i = 1,\ j = 7$	4.1934	1.8517	2.5760	2.1829
$i = 2,\ j = 3$	24.8805	17.5762	20.1018	14.8597
$i = 2,\ j = 4$	14.5753	13.4764	11.8478	10.9713
$i = 2,\ j = 5$	5.8710	0.0106	0.9531	0.2117
$i = 2,\ j = 6$	9.3098	5.4028	8.5179	6.8794
$i = 2,\ j = 7$	10.5669	10.2796	9.1351	9.0171
$i = 3,\ j = 4$	15.9624	16.1637	15.1953	15.9604
$i = 3,\ j = 5$	7.0113	0.0075	0.1420	0.0453
$i = 3,\ j = 6$	4.4844	3.0946	5.0575	3.1724
$i = 3,\ j = 7$	8.0324	2.2217	4.4287	3.2240
$i = 4,\ j = 5$	16.3438	0.0637	0.9827	0.4403
$i = 4,\ j = 6$	15.7964	10.3994	14.7998	12.6494
$i = 4,\ j = 7$	8.2665	7.5554	7.6485	7.6740
$i = 5,\ j = 6$	12.0868	0.0018	0.0506	0.0393
$i = 5,\ j = 7$	6.3027	0.0646	1.2451	1.0585
$i = 6,\ j = 7$	10.5326	7.0662	10.5267	10.2798

Table 8.13: Shares: $Q(i, j; 5)$ of the Principal Component Scores $\boldsymbol{x_t}$ of the O-SV Model, modelling $\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$ as ARMA

$\boldsymbol{B}$ $\boldsymbol{W}$	$\boldsymbol{I}$ $\boldsymbol{I}$	$\boldsymbol{I}$ $(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ $\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ $(\boldsymbol{\Lambda})^{-\frac{1}{2}}$
$i = 1,\ j = 1$	25.8247	0.0390	0.0096	0.1143
$i = 2,\ j = 2$	12.1216	0.0003	0.0037	0.0039
$i = 3,\ j = 3$	21.7717	0.0174	0.0783	0.0424
$i = 4,\ j = 4$	9.8676	0.0305	0.0386	0.0254
$i = 5,\ j = 5$	7.0357	7.8596	7.3699	7.7908
$i = 6,\ j = 6$	5.8021	12.0784	8.7190	12.7006
$i = 7,\ j = 7$	29.3466	24.5448	32.8731	22.3416
$i = 1,\ j = 2$	15.7636	0.0008	0.0030	0.0089
$i = 1,\ j = 3$	15.4336	0.0150	0.0064	0.0186
$i = 1,\ j = 4$	2.8543	0.0342	0.0141	0.0350
$i = 1,\ j = 5$	6.4805	4.7960	3.8954	4.9728
$i = 1,\ j = 6$	10.0396	6.3092	6.0085	7.2401
$i = 1,\ j = 7$	2.5720	0.6250	0.5457	0.5941
$i = 2,\ j = 3$	1.9301	0.0001	0.0031	0.0026
$i = 2,\ j = 4$	4.5543	0.0007	0.0041	0.0034
$i = 2,\ j = 5$	1.3716	0.2081	1.0117	0.5886
$i = 2,\ j = 6$	8.2485	0.0299	0.5890	0.2583
$i = 2,\ j = 7$	1.7404	0.0006	0.1362	0.0506
$i = 3,\ j = 4$	2.9803	0.0277	0.0603	0.0410
$i = 3,\ j = 5$	1.8076	1.7667	1.9711	1.6744
$i = 3,\ j = 6$	10.2044	0.9495	1.5455	0.7525
$i = 3,\ j = 7$	8.5330	1.4048	3.4477	1.1850
$i = 4,\ j = 5$	5.3189	4.1200	9.0551	5.4544
$i = 4,\ j = 6$	4.3309	1.7600	6.0638	3.1209
$i = 4,\ j = 7$	1.7243	0.8931	4.0435	1.2659
$i = 5,\ j = 6$	37.5121	31.6437	34.1975	31.1600
$i = 5,\ j = 7$	4.6273	5.4763	4.5344	5.0269
$i = 6,\ j = 7$	30.2796	50.7645	30.0812	52.7076

Table 8.14: Shares: $Q(i, j; 5)$ of the Principal Component Scores $\boldsymbol{x_t}$ of the O-SV Model, modelling $\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$ as VAR

$\boldsymbol{B}$ $\boldsymbol{W}$	$\boldsymbol{I}$ $\boldsymbol{I}$	$\boldsymbol{I}$ $(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ $\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ $(\boldsymbol{\Lambda})^{-\frac{1}{2}}$
$i=1,\ j=1$	28.3775	0.0753	0.0166	0.2088
$i=2,\ j=2$	12.1355	0.0003	0.0048	0.0050
$i=3,\ j=3$	27.1696	0.5318	0.5925	0.3287
$i=4,\ j=4$	9.8328	0.0055	0.0091	0.0060
$i=5,\ j=5$	5.8082	6.5829	6.0737	6.4741
$i=6,\ j=6$	3.5144	7.5226	4.0476	7.8868
$i=7,\ j=7$	20.5046	19.5084	24.2100	18.4059
$i=1,\ j=2$	13.9358	0.0010	0.0047	0.0138
$i=1,\ j=3$	13.7789	0.0807	0.0172	0.0515
$i=1,\ j=4$	2.1440	0.0202	0.0051	0.0128
$i=1,\ j=5$	4.8458	2.7642	2.2090	2.5980
$i=1,\ j=6$	12.4469	11.8738	10.2549	12.0810
$i=1,\ j=7$	2.9377	1.2600	1.0942	1.1691
$i=2,\ j=3$	5.3758	0.0001	0.0081	0.0068
$i=2,\ j=4$	4.2713	0.0003	0.0015	0.0012
$i=2,\ j=5$	1.6056	0.3786	1.4677	0.8798
$i=2,\ j=6$	8.9156	0.0388	0.9305	0.4244
$i=2,\ j=7$	2.8754	0.0018	0.1632	0.0662
$i=3,\ j=4$	3.0549	0.0416	0.0363	0.0248
$i=3,\ j=5$	1.7187	1.5633	1.0397	0.9481
$i=3,\ j=6$	7.4326	0.9171	2.2012	0.7374
$i=3,\ j=7$	6.0032	4.2328	4.3085	1.6545
$i=4,\ j=5$	4.1285	3.6162	9.1277	5.1341
$i=4,\ j=6$	1.1592	0.8445	2.9107	1.2559
$i=4,\ j=7$	1.3644	0.4457	2.1652	0.6244
$i=5,\ j=6$	30.4880	27.7423	30.4366	28.1500
$i=5,\ j=7$	4.1964	3.6396	3.8999	3.6547
$i=6,\ j=7$	21.8716	37.9931	20.0499	38.3199

Table 8.15: Exchange Rates: $Q(i, j; 5)$ of the Sample Residuals $\hat{\boldsymbol{e}}_t$ of the O-SV Model

$\hat{\boldsymbol{\mu}}_t$	ARMA	ARMA	ARMA	ARMA	VAR	VAR	VAR	VAR
$\boldsymbol{B}$	$\boldsymbol{I}$	$\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	$\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$
$\boldsymbol{W}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$
$i = 1, j = 1$	1.7036	1.9194	1.9541	1.9570	1.9032	2.0817	2.0842	2.0296
$i = 2, j = 2$	2.2030	2.1226	1.9050	1.9381	2.1218	2.2475	1.9665	2.0125
$i = 3, j = 3$	7.1309	6.7654	8.5710	7.4720	5.9774	5.3557	7.0663	5.5247
$i = 4, j = 4$	4.1198	0.0687	1.2476	0.1282	4.6022	0.0240	0.8085	0.0449
$i = 5, j = 5$	13.4585	17.6498	20.2911	21.3844	12.1332	16.9974	18.2835	19.5475
$i = 1, j = 2$	0.8733	0.7618	0.7189	0.6890	1.1824	1.1506	1.1289	1.0950
$i = 1, j = 3$	1.9530	1.7581	1.5273	1.4449	1.3282	1.2739	1.1014	1.1210
$i = 1, j = 4$	2.1811	1.7959	1.8482	2.0295	1.9355	1.2089	1.4830	1.3934
$i = 1, j = 5$	0.5918	0.6170	0.9950	0.9222	0.7800	0.9155	1.1183	1.0774
$i = 2, j = 3$	2.5613	2.5102	2.1804	2.1998	1.8933	1.7903	1.6182	1.5691
$i = 2, j = 4$	1.5428	1.4401	1.3094	1.2902	1.5802	1.5163	1.2293	1.2120
$i = 2, j = 5$	0.5878	0.8065	1.0648	1.1696	0.6557	0.9939	1.1457	1.2973
$i = 3, j = 4$	4.8957	2.5292	4.6361	2.3817	3.3674	1.5470	3.2274	1.3709
$i = 3, j = 5$	4.7016	4.7152	6.0992	5.9435	4.0935	4.0148	5.2714	4.9659
$i = 4, j = 5$	2.9762	2.3890	2.9930	2.5304	3.4583	2.5635	3.5166	2.8785

Table 8.16: Exchange Rates: $Q(i, j; 5)$ of the Principal Component Scores $\boldsymbol{x_t}$ of the O-SV Model

$\hat{\boldsymbol{\mu}}_t$	ARMA	ARMA	ARMA	ARMA	VAR	VAR	VAR	VAR
$\boldsymbol{B}$	$\boldsymbol{I}$	$\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	$\boldsymbol{I}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$
$\boldsymbol{W}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$
$i = 1,\ j = 1$	1.8383	98.1126	2.0393	19.1867	2.2070	41.5736	8.1040	36.5320
$i = 2,\ j = 2$	2.4396	0.1725	0.8741	0.1987	3.7994	0.2053	1.0104	0.2237
$i = 3,\ j = 3$	2.9866	3.8003	3.3586	3.0624	4.0459	3.4795	2.5501	2.4346
$i = 4,\ j = 4$	1.8944	0.8526	1.0586	0.0787	2.7417	2.1635	1.1905	0.2540
$i = 5,\ j = 5$	6.4058	5.0904	3.4249	2.9532	5.4763	4.4539	3.3865	2.9273
$i = 1,\ j = 2$	45.7996	1.9091	0.0418	0.0866	59.2947	5.9096	0.1397	0.3538
$i = 1,\ j = 3$	24.6136	36.9333	23.5519	23.4795	20.2397	41.9035	21.4190	20.7474
$i = 1,\ j = 4$	52.7449	111.0083	122.4791	155.1380	67.0461	79.6804	91.5027	119.7357
$i = 1,\ j = 5$	6.5816	12.3711	10.0861	7.2599	8.2912	9.1292	16.0179	12.1649
$i = 2,\ j = 3$	20.4133	2.1394	18.1460	8.9818	24.5563	1.2306	20.2654	6.0436
$i = 2,\ j = 4$	7.6597	3.2113	14.1136	6.8345	9.5921	5.2698	12.0329	7.4956
$i = 2,\ j = 5$	7.9477	2.7026	6.3302	3.5874	5.2808	2.6137	3.3984	2.9673
$i = 3,\ j = 4$	11.0064	13.5556	13.1783	15.0053	8.1797	11.4886	7.7356	10.1797
$i = 3,\ j = 5$	23.1558	17.7985	47.6699	44.8503	20.4908	11.7541	43.4398	37.3818
$i = 4,\ j = 5$	37.7675	34.6287	30.0288	27.8742	35.4072	32.9614	34.7459	33.8133

can give an indication of whether the non zero conditional cross correlations have a negative affect on the model results.

Firstly the choices of conditional mean model, $\boldsymbol{B}$ and $\boldsymbol{W}$ for each of the models are considered. As mentioned, this is achieved by fixing two of the variations and determining the affect of changing the third. However, the Q-statistics are each fitted to the sample residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ and the principal component scores $\boldsymbol{x}_{\boldsymbol{t}}$. It is expected that if the Q-statistics fitted to the sample residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ indicate that one of the choices if preferable to the other then the Q-statistics fitted to the principal component scores $\boldsymbol{x}_{\boldsymbol{t}}$ will indicate that the same choice is preferable. However it appears that the Q-statistics fitted to the principal component scores $\boldsymbol{x}_{\boldsymbol{t}}$ often give conflicting results to those fitted to the sample residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ in terms of which variation is best within each model.

Therefore the comments made here are for the Q-statistics fitted to the sample residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ and not $\boldsymbol{x}_{\boldsymbol{t}}$ because ultimately it is the conditional covariance of $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ that this study is interested in modelling. It is difficult to determine which of the two conditional mean models are better. For example, the Q-statistics fitted to the sample residuals $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$ for the share data suggest that it is preferable to model the conditional mean with an ARMA model in the O-GARCH or O-SV model but that VAR is preferable in the O-EWMA model. However, for the exchange rate data is it difficult to determine which of the two conditional mean models is best. With regards to $\boldsymbol{B}$, the Q-statistics suggest that it is preferable to choose $\boldsymbol{B} = \boldsymbol{I}$ in the O-GARCH models but it is difficult to say which is preferable in the other two models. However, there are some comments which can be made regarding the choice of $\boldsymbol{W}$. The choice of $\boldsymbol{W}$ makes very little difference in the sense that the Q-statistics are very similar whether $\boldsymbol{W} = \boldsymbol{I}$ or $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ for the O-GARCH model and are identical in the O-EWMA model. On the other hand the O-SV model results suggests that it is preferable to use $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$.

Secondly the three models are compared by taking an overall look at the statistics for each model. Thus if the Q-Statistics of each of the three models are considered, then the Q-statistics of the O-SV model are generally the smallest in each of the data sets, all else being equal. In the case of the share data the model with the next smallest Q-statistics in general appears to be the O-EWMA model. However, in the case of the exchange rate data it is difficult to tell whether the O-GARCH or O-EWMA models generally have the next smallest Q-statistics. Moreover further investigation is required to determine whether the O-SV model is in general preferable to the O-EWMA and O-GARCH models.

Finally the affect on the models of non zero conditional cross correlations are examined. In the case of the exchange rate data there is definitely an indication that the non zero conditional cross correlations between the principal component scores have a negative effect on the results. This can be seen by the fact that the Q-Statistics of the principal component scores are generally larger when $i \neq j$ than when $i = j$. However, for the share data there does not appear to be an indication that the fit to the principal component scores are worse when $i \neq j$ than when $i = j$. Hence it is difficult to determine whether the fact that the factors do not have zero conditional correlations will usually have a negative effect on these models for any data set.

8.3.3.2 Multivariate Portmanteau Statistic Results

The results of the multivariate portmanteau statistics can be found in Tables 8.17 to 8.22. Tables 8.17 and 8.18 contain the statistics for the O-GARCH models, Tables 8.19 and 8.20 contain the statistics for the O-EWMA models and Tables 8.21 and 8.22 contain the statistics for the O-SV models. In each table and for each lag the smallest statistic is highlighted in pink to make it easier to determine which model is best in terms of this statistic. In other words the models corresponding to the highlighted cell provides the best fit at that number of lags according to the multivariate portmanteau statistic. Once again, the multivariate portmanteau statistics for the O-EWMA models, Tables 8.19 and 8.20, do not contain the variation $\boldsymbol{W}$ as either choice of this variation gives the same results in the O-EWMA model.

Although there is no known distribution which can be used to determine the tail probabilities for observing such multivariate portmanteau statistics, it is clear that the larger the statistic the worse the model fit. By comparing two statistics at the same number of lags where the two models are identical except for the one aspect of the model which is being considered, one can get an idea of which of the two are preferable. To begin with the three variations within each model are compared, that is the choice of the conditional mean model, the choice of $\boldsymbol{B}$ and lastly the choice of $\boldsymbol{W}$. They are compared by fixing two of the variations and determining the affect of changing the third.

The results indicate that it is generally best to model the conditional mean using a VAR model. The only instance where it appears not to be the case is where the exchange rate data are modelled using an O-SV model.

Table 8.17: Shares - Multivariate Portmanteau Statistics of the O-GARCH Residuals

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	Lags				
			1	2	3	4	5
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	2578.8075	4958.0913	7190.8501	9382.5736	11560.1112
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	2579.6972	4959.9123	7193.7789	9386.7680	11565.7688
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	2604.7067	4964.0564	7180.9373	9352.3235	11501.2079
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	2604.7085	4964.0550	7180.9335	9352.3171	11501.2005
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	2520.4708	4848.7239	7048.8940	9231.1483	11406.3139
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	2521.8377	4851.4199	7053.1182	9236.9301	11413.5781
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	2559.5490	4840.9130	6987.9663	9113.6168	11210.2063
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	2559.5451	4840.9051	6987.9547	9113.6031	11210.1892

Table 8.18: Exchange Rates - Multivariate Portmanteau Statistics of the O-GARCH Residuals

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	Lags				
			1	2	3	4	5
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	1475.1832	2860.5848	4296.7231	5720.7219	7104.3950
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	1469.7089	2847.5290	4274.5692	5693.2942	7066.4075
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	1471.2101	2850.6252	4270.4248	5683.9320	7073.2393
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	1471.2072	2850.6212	4270.4213	5683.9284	7073.2358
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	1452.8505	2826.2424	4266.0373	5647.8986	7009.5924
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	1447.1411	2812.8907	4245.0798	5621.1975	6975.7647
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	1466.4356	2864.1538	4324.0659	5733.9186	7143.2130
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	1453.6079	2816.5245	4239.2862	5614.1579	6979.8291

Table 8.19: Shares - Multivariate Portmanteau Statistics of the O-EWMA Residuals

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$	$\boldsymbol{B}$	Lags				
		1	2	3	4	5
ARMA	$\boldsymbol{I}$	2087.6761	3963.1160	5703.6890	7430.1481	9110.0407
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	2117.6298	3965.5730	5690.4956	7388.3150	9040.2540
VAR	$\boldsymbol{I}$	2045.5424	3872.8672	5569.5052	7299.4923	8965.5091
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	2077.7861	3885.2939	5566.4589	7266.2481	8924.7864

Table 8.20: Exchange Rates - Multivariate Portmanteau Statistics of the O-EWMA Residuals

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$	$\boldsymbol{B}$	Lags				
		1	2	3	4	5
ARMA	$\boldsymbol{I}$	1181.7002	2226.1734	3271.1599	4362.5923	5408.5681
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	1185.7125	2245.8589	3290.1992	4391.3465	5464.8077
VAR	$\boldsymbol{I}$	1167.4307	2200.1931	3254.2217	4281.9862	5292.8147
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	1179.5959	2219.6644	3275.9097	4320.7168	5354.5109

Table 8.21: Shares - Multivariate Portmanteau Statistics of the O-SV Residuals

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	Lags				
			1	2	3	4	5
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	2285.9702	4218.7039	6007.7523	7808.2706	9531.3594
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	2191.6345	4024.7429	5671.5918	7519.1518	9649.0560
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	2240.7644	4113.6660	5783.2225	7607.3608	9785.7275
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	2233.1308	4090.6274	5743.7601	7552.8878	9731.0470
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	2226.0971	4101.9715	5830.1099	7613.0776	9317.0726
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	2142.2958	3959.2292	5573.0397	7366.8980	9500.1291
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	2157.8471	3983.9716	5593.5462	7350.9038	9495.9937
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	2159.1249	3977.2349	5577.7188	7318.6287	9454.9227

Table 8.22: Exchange Rates - Multivariate Portmanteau Statistics of the O-SV Residuals

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	Lags				
			1	2	3	4	5
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	1095.6546	2062.9205	3039.2601	4032.5414	5001.7891
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	2377.7767	3301.8940	4208.7053	5093.7984	5967.7731
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	2256.3267	3223.1646	4172.1998	5049.5537	5895.4841
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	2274.8945	3246.1086	4205.3983	5084.6669	5954.4346
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	1072.1112	2029.7348	3021.3097	3970.3691	4899.7663
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	2602.5932	3525.2381	4422.1194	5267.2495	6112.0510
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	2486.8039	3430.3887	4369.0415	5222.3006	6050.6342
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	2515.8199	3478.7045	4434.0383	5286.9005	6140.4728

On the other hand, there does not appear to be a clear indication as to whether it is preferable to use $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ or for $\boldsymbol{B} = \boldsymbol{I}$.

Lastly the two choices of $\boldsymbol{W}$ are compared. In the O-GARCH model for the share data it appears best to choose $\boldsymbol{W} = \boldsymbol{I}$ when $\boldsymbol{B} = \boldsymbol{I}$ and very little difference in the choice of $\boldsymbol{W}$ when $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ but for the exchange rate data choosing $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ is preferable. As previously mentioned, the O-EWMA model will always give identical results whether $\boldsymbol{W} = \boldsymbol{I}$ or whether $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$. This is in contrast with the O-SV model which gives opposite results for the two datasets. It seems best to choose $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ when modelling the share data but it appear best to choose $\boldsymbol{W} = \boldsymbol{I}$ when modelling the exchange rate data. Hence for this aspect of the model there does not appear to be any consistency throughout.

The best combination of the three model variations for the O-GARCH, O-EMWA and O-SV models in terms of the Multivariate Portmanteau statistic can be determined by identifying which combination of variations usually has the smallest statistic. In the O-GARCH model it appears that using a VAR model and $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ usually give the smallest statistics for both data sets. Additionally $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ appears best in the share data set but there is no clear indication of which of the two possible matrices for $\boldsymbol{B}$ is better for the O-GARCH models. In the O-EWMA model using a VAR model also appears to be preferable for both data sets but with regards to $\boldsymbol{B}$, there is no clear indication of which choice of is best for the share data and the exchange rate data suggests $\boldsymbol{B} = \boldsymbol{I}$ is preferable. Recall that either choice of $\boldsymbol{W}$ give the same results for the O-EWMA model. Lastly when fitting the O-SV model to either data set using a VAR model is best and $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ is preferable for the share data but $\boldsymbol{W} = \boldsymbol{I}$ is preferable for the exchange rate data. There is no clear indication for the share data as to which choice of $\boldsymbol{B}$ is preferable but for the exchange rate data $\boldsymbol{B} = \boldsymbol{I}$ is preferable.

Finally the O-GARCH, O-EWMA and O-SV models are compared to one another. This comparison can be done in two ways. The first is to compare the three models keeping all else equal and determine which of the models on average has the smallest statistic. If this method is used then the O-EWMA model in general seems to be best for both datasets except when $\boldsymbol{B} = \boldsymbol{I}$ and $\boldsymbol{W} = \boldsymbol{I}$ where the O-SV model is best. The second method is to compare the smallest statistic at each lag for the three models within each data set. In this case the O-EWMA is best followed by the O-SV model for the share data and in the case of the exchange rate data O-SV is best followed by the O-EWMA model. However, the second method may in fact be more suitable as in practice one would choose the combination of variations which is best for that specific model and then compare the O-GARCH, O-EWMA and O-SV models using the best variation for each.

8.3.3.3 AMAD Statistic Results

The AMAD statistics are presented in in Tables 8.23 to 8.28. Tables 8.23 to 8.24 contain the AMAD statistics for the O-GARCH models, Tables 8.25 to 8.26 for the O-EWMA models and Tables 8.27 and 8.28 for the O-SV models. Similar to the multivariate portmanteau tables, the smallest AMAD statistic in each column is highlighted in pink.

The AMAD statistics can be used to make the same comparisons as the other two statistics, that is the comparison of the three models and the choice of conditional mean model, $\boldsymbol{B}$ and $\boldsymbol{W}$. However in addition to this the AMAD statistic has other purposes. These are that the AMAD statistics can help to determine the optimal number of time points to use to estimate the model parameters, whether the estimate is best for one time point or for the region around that time point and how the forecast accuracy is affected by the number of time steps into the future the forecast is taken. However the comparisons are only for the data sets in the present study and cannot necessarily be generalised. Further investigation would be required to prove that the results are in general true.

Firstly the three model variations are discussed, that is the choice of conditional mean model, $\boldsymbol{B}$ and $\boldsymbol{W}$. Modelling the conditional mean with a VAR model generally results in a smaller AMAD statistic than using an ARMA model, while keeping all else equal. The next variation which is compared is the choice of $\boldsymbol{B}$. For the O-GARCH and O-EWMA models choosing $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ is better than choosing $\boldsymbol{B} = \boldsymbol{I}$. On the other hand, in the O-SV models it appears that choosing $\boldsymbol{B} = \boldsymbol{I}$ is preferable when $\boldsymbol{W} = \boldsymbol{I}$ but there appears to be little difference when $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$. The last variation discussed is the choice of $\boldsymbol{W}$. When fitting an O-GARCH model to the share data it appears preferable to choose $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ but when fitting an O-GARCH model to the exchange rate data it is unclear which is preferable. However when fitting an O-SV model to either of the data sets it appears preferable to choose $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$. In contrast, either choice of $\boldsymbol{W}$ give identical results for the O-EWMA models.

Table 8.23: Shares: AMAD Statistic of O-GARCH

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	$v=0$ $T_A=500$ $c=1$	$v=0$ $T_A=500$ $c=5$	$v=0$ $T_A=1000$ $c=1$	$v=0$ $T_A=1000$ $c=5$	$v=1$ $T_A=500$ $c=1$	$v=1$ $T_A=500$ $c=5$	$v=1$ $T_A=1000$ $c=1$	$v=1$ $T_A=1000$ $c=5$
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	54.4997	53.9323	54.3268	53.8473	34.8015	36.1013	34.8736	36.1237
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	54.1668	53.5809	54.0105	53.5350	34.5845	35.8902	34.6722	35.9172
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	54.2765	53.7618	54.1707	53.6411	34.6020	35.9943	34.6754	35.9595
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	53.9718	53.4579	53.8391	53.3371	34.4102	35.8105	34.4643	35.7715
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	53.8708	53.4220	53.7198	53.3009	34.4221	35.8728	34.4360	35.8025
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	53.8612	53.4006	53.7117	53.3045	34.4151	35.8629	34.4261	35.7986
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	53.6690	53.2720	53.4847	53.0434	34.1970	35.7944	34.1511	35.5901
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	53.6662	53.2336	53.4851	53.0435	34.2118	35.7543	34.1522	35.5891

Table 8.24: Exchange Rates: AMAD Statistic of O-GARCH

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	$v=0$ $T_A=500$ $c=1$	$v=0$ $T_A=500$ $c=5$	$v=0$ $T_A=1000$ $c=1$	$v=0$ $T_A=1000$ $c=5$	$v=1$ $T_A=500$ $c=1$	$v=1$ $T_A=500$ $c=5$	$v=1$ $T_A=1000$ $c=1$	$v=1$ $T_A=1000$ $c=5$
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	17.9671	18.1432	17.9320	18.1228	11.8675	13.6279	12.2891	13.6580
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	17.9673	18.1428	17.9333	18.1128	11.8671	13.6281	12.2896	13.6465
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	17.8968	18.2127	17.6810	17.9262	11.9110	13.7143	12.1864	13.5264
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	17.9069	18.1774	17.6783	17.9259	11.9172	13.6831	12.1842	13.5262
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	16.5376	16.2741	16.4121	16.2524	10.6120	11.8181	10.9373	11.8755
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	16.5366	16.2727	16.4121	16.2525	10.6132	11.8169	10.9372	11.8759
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	16.4580	16.1667	16.1502	16.0512	10.6080	11.7078	10.7974	11.7143
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	16.4417	16.2092	16.1508	16.0513	10.5946	11.7500	10.7971	11.7143

Table 8.25: Shares: AMAD Statistic of O-EWMA

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$	$\boldsymbol{B}$	$v=0$ $T_A=500$ $c=1$	$v=0$ $T_A=500$ $c=5$	$v=0$ $T_A=1000$ $c=1$	$v=0$ $T_A=1000$ $c=5$	$v=1$ $T_A=500$ $c=1$	$v=1$ $T_A=500$ $c=5$	$v=1$ $T_A=1000$ $c=1$	$v=1$ $T_A=1000$ $c=5$
ARMA	$\boldsymbol{I}$	55.0128	55.0979	55.1113	55.1185	35.7258	36.9134	35.5901	36.8945
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	54.6402	54.7578	55.0011	55.1381	35.6194	36.7013	35.5093	36.9318
VAR	$\boldsymbol{I}$	54.1183	54.2160	54.2509	54.2766	35.3161	36.4886	35.0725	36.3437
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	53.7650	53.8915	54.1647	54.2490	35.1735	36.2784	34.8956	36.3060

Table 8.26: Exchange Rates: AMAD Statistic of O-EWMA

$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}$	$\boldsymbol{B}$	$v=0$ $T_A=500$ $c=1$	$v=0$ $T_A=500$ $c=5$	$v=0$ $T_A=1000$ $c=1$	$v=0$ $T_A=1000$ $c=5$	$v=1$ $T_A=500$ $c=1$	$v=1$ $T_A=500$ $c=5$	$v=1$ $T_A=1000$ $c=1$	$v=1$ $T_A=1000$ $c=5$
ARMA	$\boldsymbol{I}$	19.4421	19.8228	19.3488	19.8350	14.9603	15.4773	14.7569	15.4774
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	18.6114	18.9101	19.1685	19.6302	14.0532	14.4125	14.7066	15.3408
VAR	$\boldsymbol{I}$	17.0956	17.2340	17.5206	17.6931	12.4705	13.0267	12.8578	13.3328
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	16.5149	16.7297	17.1484	17.3610	12.1931	12.4886	12.8006	13.2520

Table 8.27: Shares: AMAD Statistic of O-SV

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	$v=0$ $T_A=500$ $c=1$	$v=0$ $T_A=500$ $c=1$	$v=0$ $T_A=1000$ $c=1$	$v=0$ $T_A=1000$ $c=1$	$v=1$ $T_A=500$ $c=1$	$v=1$ $T_A=500$ $c=1$	$v=1$ $T_A=1000$ $c=1$	$v=1$ $T_A=1000$ $c=1$
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	52.7655	52.7654	52.7655	52.7655	43.8372	43.8371	43.8372	43.8371
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	51.8729	51.9761	51.4296	51.4303	39.5529	39.4947	40.2843	40.0795
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	56.4025	56.3437	54.0335	54.0136	45.9145	45.7895	43.9934	43.9148
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	51.9798	52.0438	51.4212	51.3886	39.5996	39.4806	40.2662	40.0085
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	51.5689	51.5689	51.5690	51.5689	43.0655	43.0654	43.0655	43.0654
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	50.9045	51.0153	50.3003	50.3162	38.9035	38.8495	39.4966	39.2999
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	55.5785	55.5072	53.0884	53.0683	45.4673	45.3394	43.4397	43.3588
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	50.9329	51.0017	50.2768	50.2549	38.9115	38.7759	39.4798	39.2137

Table 8.28: Exchange Rates: AMAD Statistic of O-SV

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	$\boldsymbol{W}$	$v=0$ $T_A=500$ $c=1$	$v=0$ $T_A=500$ $c=1$	$v=0$ $T_A=1000$ $c=1$	$v=0$ $T_A=1000$ $c=1$	$v=1$ $T_A=500$ $c=1$	$v=1$ $T_A=500$ $c=1$	$v=1$ $T_A=1000$ $c=1$	$v=1$ $T_A=1000$ $c=1$
ARMA	$\boldsymbol{I}$	$\boldsymbol{I}$	17.3394	17.3386	17.3394	17.3391	16.1923	16.1912	16.1924	16.1915
ARMA	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	17.6716	17.5943	16.9869	16.9287	14.3189	13.8614	14.2319	13.9096
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	32.4082	32.1928	23.4752	23.4108	30.5503	30.3907	21.6710	21.6127
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	17.7169	17.6300	17.0421	16.9857	14.3499	13.8895	14.2311	13.8668
VAR	$\boldsymbol{I}$	$\boldsymbol{I}$	16.1486	16.1485	16.1487	16.1488	14.9913	14.9905	14.9919	14.9910
VAR	$\boldsymbol{I}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	15.7682	15.7170	15.4525	15.4267	12.6522	12.3447	12.8661	12.6289
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$\boldsymbol{I}$	31.6468	31.4211	22.4408	22.3712	29.9105	29.7233	20.7026	20.6387
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	$(\boldsymbol{\Lambda})^{-\frac{1}{2}}$	15.7039	15.6745	15.4542	15.4330	12.5664	12.2881	12.8240	12.5730

Therefore to obtain an idea of which model variation within the O-GARCH, O-EWMA and O-SV models is best, at least in terms of the AMAD statistic, the smallest statistic in each column of the tables is considered. For both the O-GARCH and O-EWMA models it appears best to model the conditional mean with a VAR model and $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$. However the best choice of $\boldsymbol{W}$ is unclear in the O-GARCH model and irrelevant in the O-EWMA model. Lastly, the AMAD statistics fitted to the O-SV model gives the impression that it is best to model the conditional mean using VAR and $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ but it is unclear which choice of $\boldsymbol{B}$ is better.

Therefore to get a sense of whether the O-GARCH, O-EWMA or O-SV model is best the smallest AMAD statistic in each data set with the same values of T_A, v and c is considered. Recall that T_A is the number of time points used to estimate the parameters in the statistic, v is the number of adjacent residuals on each side to include in the proxy and c refers to the number of time points into the future the forecast is. Using this method it would appear that the O-SV model is best when $v = 0$ but O-GARCH is best when $v = 1$. This suggests that the O-SV model gives the best forecast of a point in the future but O-GARCH is best at estimating the covariance in a small region around the point forecasted.

The third set of comparisons are the number of time points used to estimate the model parameters. Hence to determine whether it is best to use 500 or 1000 time points the statistics of each must be compared with all else equal. In other words, if the statistics using 500 time points are the smaller of the two for each of the model variations and values of c and v, then it is generally best to estimate the parameters using 500 time points and visa versa. Of course this is likely to depend on the specific data set as well as the particular model, that is O-GARCH, O-EWMA and O-SV. Therefore in the case of the O-GARCH and O-SV models, $T_A = 1000$ is generally better than $T_A = 500$ when $v = 0$. Hence it appears that using a longer time period to estimate the parameters is preferable. However when $v = 1$ there appears to be no clear indication as to whether 500 or 1000 data points are preferable. On the other hand, in case of the O-EWMA models it would seem that $T_A = 500$ is generally better than $T_A = 1000$. The only exception is the case where the O-EMWA model fitted to the share data and $v = 0$.

The fourth comparison relates to the choice of v. Recall that the v is the number of adjacent residuals on each side included in the proxy of the conditional covariance. The AMAD statistic of any two identical models is smaller when $v = 1$ than when $v = 0$. This indicates that the forecasts of the orthogonal models give a better estimate of the volatility of the data around that time point as opposed to for that point only. Although not done in this study, it is possible to determine the so called optimal vicinity (that is the value of v) by testing a wider range of the values of v to determine the time period over which the volatility provides the best estimates.

The fifth and final point is how the accuracy of the model forecast is affected by the number of time points into the future the forecast is made. It is intuitive that $c = 1$ would be better than $c = 5$ as in most models the further into the future an estimate is made the more likely it is to be inaccurate. However, the results show that while this is true for some of the models tested in this thesis it does not always appear to be the case. For the O-GARCH model $c = 1$ is preferable except for the case where the O-GARCH model is fitted to the share data and $v = 1$. Similarly, in the case of the O-EWMA models, $c = 1$ generally results in a smaller AMAD statistic than $c = 5$. Therefore the results of the O-GARCH and O-EWMA models are generally in line with what one would expect to see. On the other hand, the O-SV models fitted to the exchange rate returns give smaller AMAD statistics for $c = 5$ than for $c = 1$. In addition, the AMAD statistics for the O-SV model fitted to the share data are generally smaller for $c = 5$ when $v = 1$. Hence this appears to contradict what intuition suggests.

8.4 Collation of the Results

The results described in Section 8.2 pertaining to the plots and Section 8.3 relating to the three statistics for goodness of fit are now collated. Thus the model variations, the three models and the effect of non zero conditional correlations on the results are reviewed in that order.

Firstly results of the model variations, that is the conditional mean model, $\boldsymbol{B}$ and $\boldsymbol{W}$, are collated. Although graphically it is difficult to determine the differences in the conditional covariance estimates for different variations, the differences are more easily observed in the statistics. Thus these variations are considered solely in terms of the goodness of fit statistics. The multivariate portmanteau statistic and AMAD statistics both indicate that the conditional mean is best modelled using a VAR model as opposed to an ARMA model. However the Q-Statistics do not give a clear indication whether a VAR or ARMA model is preferable.

On the other hand there does not seem to be any agreement between the various statistics with regard to whether $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$ or $\boldsymbol{B} = \boldsymbol{I}$ is preferable. Even for just one statistic, there is often no agreement as to which of the two is better for a given model.

The third and final variation reviewed is $\boldsymbol{W}$. The statistics for the O-GARCH model are often similar whether $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ or $\boldsymbol{W} = \boldsymbol{I}$ with all else equal. However the O-EWMA model is identical for either choice of $\boldsymbol{W}$. It appears that the three statistics suggest that it is generally preferable to model $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ in the O-SV model rather than $\boldsymbol{W} = \boldsymbol{I}$. Thus Bongers (2008) suggestion of using $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ does not appear to be necessary except in the case of the O-SV model.

The second area where the results are considered in aggregate are the three model comparisons. The evidence in favour of the O-GARCH model is that the AMAD statistics indicate it is best when $v = 0$ and the plots of the conditional variance and correlation estimates are reasonable however the other two statistics indicate that it is the worse out of the three models. The evidence in favour of the O-EWMA model is that the multivariate portmanteau statistics indicate that it is best in the case of the share data. However, due to the unreasonable estimates of the conditional correlations for the share data for the O-EWMA model in Figure 8.2, it is ruled out as a possiblity for this data set. Furthermore the AMAD statistic suggests that the O-EWMA model gives the worse forecasts out of the three. This could be attributed to the fact that the O-EWMA model forecasts have a constant term structure (Alexander, 2000). Finally there is much evidence in favour of the O-SV model and little against it. The Q-statistics suggest that the O-SV model is preferable for both data sets, the multivariate portmanteau statistics indicate that is is best for the exchange rate data set and the AMAD statistics indicate that the O-SV model is the best at estimating the conditional covariance when the conditional covariance proxy uses $v = 0$. Furthermore the plots of the conditional variance and correlation estimates suggest that these estimates are reasonable. Thus overall the results seems to indicate that an O-SV model is preferable and for this model it is best to use a VAR model for the conditional mean, $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ but the choice of $\boldsymbol{B}$ is not so clear.

Thirdly the appropriateness of the principal component scores method is discussed. There is no evidence that using the principal component scores as though they are conditionally uncorrelated in the orthogonal models is always unsuitable. However what can be said is that for some data sets it is definitely not suitable. Therefore further investigation is required to determine in which instances this method may be suitable and in which instances a more appropriate method of calculating the conditionally uncorrelated components are required.

Chapter 9

Conclusions

9.1 Aims Achieved

In this thesis, both the theory and the steps required to implement a general orthogonal factor model have been explored. Moreover, three specific orthogonal factor models, that is the O-GARCH, O-EWMA and O-SV models, as well as three different adjustments, that is the conditional mean, and the transformation matrices $\boldsymbol{B}$ and $\boldsymbol{W}$, have been discussed and evaluated in great detail. Hence the focus of this thesis has been on attempting to improve the results of the model by fitting various univariate conditional variance models to the factors and testing adjustments made to the data and model output. This is in contrast to the focus of many conditional covariance factor model papers which focus on different methods of calculating the factors.

Thus this thesis enables someone with little knowledge of orthogonal factor models to grasp the concept and theory behind such models. Furthermore, it provides the detailed steps necessary to facilitate easy implementation of the models.

The suitability of using principal component scores as factors have been explored primarily by means of the Q-statistics. The results suggest that the appropriateness of using principal component scores as factors depends on the specific data set and does not always have a negative impact on the model estimates. However, the contexts in which the principal component scores are suitable as factors is beyond the scope of this thesis. Thus further investigation is required to determine in which contexts they are suitable. Nonetheless, what can be observed is that the two data sets tested are very different in sense that the series in one are highly correlated and are not in the other. The method of principal component score did not appear appropriate for the highly correlated exchange rate data set. It did not appear appropriate in the sense that the Q-statistics fitted to the principal component scores $\boldsymbol{x_t}$ indicated the presence of non zero conditional cross correlations. Thus it is possible that the method of principal components is only suitable when the return series are not highly correlated. However, this is purely speculative and may not be the reason that the method is appropriate.

Additionally the graphical output indicates that all the series of principal component scores need to be included in the models, even for the highly correlated exchange rate data set. This is contrary to the suggestion of Alexander (2000) that only a small number of series need to be included when the data are highly correlated. On the other hand, it is possible that all the series need to be included in the model when only a small number of series are being modelled, although further research is necessary to conclude that this is usually the case.

With regards to the three models tested, the results seem to point to the O-SV model being best when fitted to either data set. Although, some of the results suggested that the O-EWMA model is best for the share data, the visual output of the conditional correlation estimates indicate that these estimates are not reasonable. For this reason the O-EWMA model is considered unsuitable for the share data set. Thus overall the results seems to indicate that an O-SV model is preferable and for this model it is best to use a VAR model for the conditional mean and $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ but the choice of $\boldsymbol{B}$ is not so clear.

On the other hand, there was much conflict in the evaluation of the best choice for the conditional mean, $\boldsymbol{B}$ and $\boldsymbol{W}$. Therefore only a few conclusions can be drawn in this regard. The first is that some of the results indicate that it is best to model the conditional mean with a VAR model yet in other results it was unclear. Secondly there is no agreement between the various statistics with regard to the choice of $\boldsymbol{B}$. Finally it is generally preferable to model $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ in the O-SV model but either choice of $\boldsymbol{W}$ gives the same results in the O-EWMA model and similar results in the O-GARCH model. Thus Bongers (2008) suggestion to use $\boldsymbol{W} = (\boldsymbol{\Lambda})^{-\frac{1}{2}}$ does not appear to be necessary except in the case of the O-SV model, which is not the context in which he tested its usage.

Finally the results of the two data sets are compared. It appears that overall the model fits to the exchange rate returns were better. They were better in the sense that the statistics were usually smaller and the plots of the conditional correlation estimates looked more reasonable. This is despite the fact that using the principal component scores as factors for the exchange rate data set is not suitable but appears suitable for the share data set. It is not suitable in the sense that the Q-statistics indicated that the presence of non zero conditional cross correlations have a negative impact on the results. Hence the results are contradictory in the sense that the method of principal component scores is unsuitable yet overall the results are better. A possible reason for this could be the claim made by Alexander (2000) that it is best to group highly correlation series together when modelling the data. However, from the results in this thesis it is difficult to determine whether this is valid as only two data sets have been tested, from which is clearly insufficient to draw inference and the results are contradictory.

A small extension to this thesis that could give more of an indication of the effect the correlations of the return series on the results is to split the share data into two data sets. The data should be split up so that one set contains only banking shares and the other contains the remaining shares and one of the banking shares. The results of these two can then be compared. Nonetheless, a more thorough investigation of the impact of the correlations of the returns would be required to draw general conclusions.

9.2 Extensions and Future Work

The following work is beyond the scope of this thesis:

1. To investigate the circumstances under which it is and is not appropriate to use the principal components as factors which are assumed to be conditionally uncorrelated. More specifically consider whether the correlation of the time series being modelled affects this.

2. To test orthogonal factor models to determine whether the model fit is typically better for a highly correlated data set than for a less correlated data set where the two data sets contain similar quantities.

3. To examine whether the O-SV model is usually preferable to the O-GARCH and O-EWMA models when modelling financial data and potentially investigate other univariate conditional variance models to fit to the principal component scores.

4. To determine whether it is typically best to include all series of the principal component scores in the model when the number of series being modelled is small. Similarly to investigate whether it is better, worse or there is little difference between include only a subset or all of the possible series of principal component scores in the model when a large number of highly correlated time series are being modelled.

Appendix A

Gamma and Beta Random Variables

A.1 Gamma and Beta Probability Density Functions

$Z \sim G(\alpha, \beta)$ denotes that Z is a gamma random variable with parameters α and β such that the probability density function is

$$f(z; \alpha, \beta) = \frac{z^{\alpha-1}\, \beta^{\alpha}\, exp(-z\beta)}{\Gamma(\alpha)}$$

with $x, \alpha, \beta > 0$.

$V \sim beta(\alpha, \beta)$ denotes that V is a beta random variable with parameters α and β such that the probability density function is

$$f(v; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)\, v^{\alpha-1}\, (1 - v)^{\beta-1}}{\Gamma(\alpha)\, \Gamma(\beta)}$$

with $0 < v < 1$ and $\alpha, \beta > 0$.

A.2 Product of a Constant and a Gamma and Beta Random Variable

If given the following information

$$\begin{array}{rcl} \gamma_{i,t-1}|F_{t-1} & \sim & gamma(a_{i,t-1},\, b_{i,t-1}) \\ \gamma_{it} & = & exp(r_{it})\, \gamma_{i,t-1}\, \eta_{it} \\ \eta_{it} & \sim & beta(\omega_i\, a_{i,t-1},\, (1-\omega_i)\, a_{i,t-1}) \\ r_{it} & & \text{is a constant} \\ \gamma_{i,t-1} \text{ and } \eta_{it} & & \text{are independent} \end{array}$$

then it can be proved that

$$\gamma_{it}|F_{t-1} \quad \sim \quad gamma(\omega_i\, a_{i,t-1},\, exp(-r_{it})\, b_{i,t-1}).$$

η_{it} does not depend on F_{t-1}, therefore the distribution of η_{it} and $\eta_{it}|F_{t-1}$ are the same.

The subscripts in the notation make a proof difficult to follow, therefore to simplify matters the notation of the information is simplified such that

$$\begin{array}{rcl} X & \sim & gamma(\alpha,\, \beta) \\ W & = & c\, X\, Y \\ Y & \sim & beta(\theta\, \alpha,\, (1-\theta)\, \alpha) \\ c & & \text{is a constant} \\ X \text{ and } Y & & \text{are independent} \end{array}$$

and that given this information it must be proved that

$$W \quad \sim \quad gamma(\theta\, \alpha,\, \frac{\beta}{c}).$$

Therefore in the new notation α is equivalent to $a_{i,t-1}$, β to $b_{i,t-1}$, c to $exp(r_{it})$, ω_i to θ, X to $\gamma_{i,t-1}$, Y to η_{it} and W to γ_{it}. In addition all the distributions in the new notation are unconditional whereas all the distributions in the original notation are conditional (since the distribution of η_{it} is equivalent to $\eta_{it}|F_{t-1}$). However the proof is equally valid for conditional distributions.

A.2.1 Result 1 - Product of a Gamma and Beta

X and Y are independent so their join probability density function is the product of the two marginal probability densities. Let $g_{XY}(x, y)$ be the joint probability density function of X and Y at $X = x$ and $Y = y$, which will be

$$g_{XY}(x, y) = \frac{x^{\alpha-1}\, \beta^{\alpha}\, exp(-x\beta)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha)\, y^{\theta\alpha-1}\, (1-y)^{(1-\theta)\,\alpha-1}}{\Gamma(\theta\,\alpha)\,\Gamma((1-\theta)\,\alpha)}.$$

X and Y will be transformed to new variables U and V such that

$$\begin{aligned} V &= X \\ U &= XY. \end{aligned}$$

with inverse functions

$$\begin{aligned} X &= V \\ Y &= \frac{U}{V}. \end{aligned}$$

The bounds of X and Y are

$$\begin{aligned} 0 < &\; X < \infty \\ 0 < &\; Y < 1 \end{aligned}$$

and the bounds of the two new variables U and V are

$$0 < U < V < \infty.$$

The Jacobian ($|J|$) of the transformation is

$$|J| = \begin{vmatrix} 0 & 1 \\ \frac{1}{V} & -\frac{U}{V^2} \end{vmatrix} = \frac{1}{V}.$$

Using information above, the joint distribution of U and V can be found by substituting the inverse function of u and v into x and y in $g_{XY}(x, y)$ and multiplying this quantity by the Jacobian. Let the join density of U and V be $g_{UV}(u, v)$ such that

$$\begin{aligned} g_{UV}(u, v) &= \frac{v^{\alpha-1}\, \beta^{\alpha}\, exp(-v\beta)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha)\, (\frac{u}{v})^{\theta\alpha-1}\, (1-\frac{u}{v})^{(1-\theta)\,\alpha-1}}{\Gamma(\theta\,\alpha)\,\Gamma((1-\theta)\,\alpha)} \times \frac{1}{v} \\ &= v^{\alpha-1}\, \beta^{\alpha}\, exp(-v\beta) \times \frac{(\frac{u}{v})^{\theta\alpha-1}\, (\frac{v-u}{v})^{(1-\theta)\,\alpha-1}}{\Gamma(\theta\,\alpha)\,\Gamma((1-\theta)\,\alpha)} \times \frac{1}{v} \\ &= v^{\alpha-1-\theta\alpha+1-(1-\theta)\,\alpha+1-1}\, \beta^{\alpha}\, exp(-v\beta) \times \frac{u^{\theta\alpha-1}\, (v-u)^{(1-\theta)\,\alpha-1}}{\Gamma(\theta\,\alpha)\,\Gamma((1-\theta)\,\alpha)} \\ &= \frac{\beta^{\alpha}\, exp(-v\beta)\, u^{\theta\alpha-1}\, (v-u)^{(1-\theta)\,\alpha-1}}{\Gamma(\theta\,\alpha)\,\Gamma((1-\theta)\,\alpha)}. \end{aligned}$$

The joint distribution can be used to determine the marginal distributions. The marginal probability density function of U (i.e. the product of the gamma and beta random variable) is the integral of the joint probability density function with respect to v over the bounds of v. Let $g_U(u)$ be the probability density function of U at $U = u$, which is

$$\begin{aligned} g_U(u) &= \int_u^\infty \frac{\beta^\alpha \, exp(-v\beta) \, u^{\theta\alpha-1} \, (v-u)^{(1-\theta)\,\alpha-1}}{\Gamma(\theta\,\alpha)\,\Gamma((1-\theta)\,\alpha)} \, dv \\ &= \frac{\beta^\alpha \, u^{\theta\alpha-1}}{\Gamma(\theta\,\alpha)\,\Gamma((1-\theta)\,\alpha)} \int_u^\infty exp(-v\beta)\,(v-u)^{(1-\theta)\,\alpha-1}\,dv \end{aligned}$$

For ease of integration, the variable v is transformed to z with $z = v - u$ so that $0 < z < \infty$ and the Jacobian is 1. The integral becomes

$$\begin{aligned} g_U(u) &= \frac{\beta^\alpha \, u^{\theta\alpha-1}}{\Gamma(\theta\,\alpha)\,\Gamma((1-\theta)\,\alpha)} \int_0^\infty exp(-z\beta - u\beta)\, z^{(1-\theta)\,\alpha-1}\,dz \\ &= \frac{\beta^\alpha \, u^{\theta\alpha-1} exp(-u\beta)\,\Gamma((1-\theta)\alpha)}{\Gamma(\theta\,\alpha)\,\Gamma((1-\theta)\,\alpha)\,\beta^{(1-\theta)\alpha}} \int_0^\infty \frac{\beta^{(1-\theta)\alpha} exp(-z\beta)\,z^{(1-\theta)\,\alpha-1}}{\Gamma((1-\theta)\alpha)}\,dz \\ &= \frac{\beta^\alpha \, u^{\theta\alpha-1} exp(-u\beta)\,\Gamma((1-\theta)\alpha)}{\Gamma(\theta\,\alpha)\,\Gamma((1-\theta)\,\alpha)\,\beta^{(1-\theta)\alpha}} \\ &= \frac{\beta^\alpha \, u^{\theta\alpha-1} exp(-u\beta)}{\Gamma(\theta\,\alpha)\,\beta^{(1-\theta)\alpha}} \\ &= \frac{\beta^{\theta\alpha} \, u^{\theta\alpha-1} exp(-u\beta)}{\Gamma(\theta\,\alpha)}. \end{aligned}$$

Hence $U \sim gamma(\theta\alpha, \beta)$ where U=XY, $X \sim gamma(\alpha, \beta)$ and $Y \sim beta(\theta\,\alpha, (1-\theta)\,\alpha)$.

A.2.2 Result 2 - Product of a Gamma and a Constant

Let c be a constant, $U \sim gamma(\theta\alpha, \beta)$ and $W = cXY = cU$. Therefore the density of W can be found by substituting the inverse function of w (i.e. $u = \frac{w}{c}$) into u in $g_U(u)$ and multiplying this by the Jacobian, which is $\frac{1}{c}$. Let

$g_W(w)$ be the probability density of W at $W = w$ which is

$$\begin{aligned} g_W(w) &= \frac{\beta^{\theta\alpha}\,(\frac{w}{c})^{\theta\alpha-1} exp(-(\frac{w}{c})\beta)}{\Gamma(\theta\,\alpha)} \times \frac{1}{c} \\ g_W(w) &= \frac{(\frac{\beta}{c})^{\theta\alpha}\,w^{\theta\alpha-1} exp(-w\,(\frac{\beta}{c}))}{\Gamma(\theta\,\alpha)}. \end{aligned}$$

Therefore

$$W \sim gamma(\theta\,\alpha,\, \frac{\beta}{c})$$

which proves what was required.

Appendix B

ARMA(p,q) and VAR(6) Parameters and Probabilities

This appendix contains the estimates of the ARMA(p,q) model parameters where the values of p and q are the optimal values for each share and exchange rate up to a maximum of 6. It also includes the estimated parameters of the VAR(6) models fitted to each of the data sets.

A t-statistic can be calculated for each estimated parameter. There is a probability associated with each t-statistic which is the probability of observing that parameter estimate or a more extreme estimate under the assumption that the true parameter is zero. However because this is a two sided test the probability required is in fact $2Pr(t > |t_{\text{observed}}|)$. If this probability is fairly small, for example less than 0.05, then the null hypothesis is rejected so that the actual parameter is assumed to be non zero so it should be included in the model.

B.1 Shares - ARMA(p,q)

B.1.1 Parameters

Table B.1: Shares - Estimates of the constants and AR parameters

	c	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
ABSA	3.03e-04	1.22	-0.73	0.78	-0.68	-	-
FirstRand	4.89e-04	0.34	0.11	-0.42	-0.62	0.67	-0.09
Std Bank	8.59e-04	0.04	-0.21	0.01	0.59	-0.45	-
Nedbank	4.52e-05	1.83	-1.93	0.88	-0.15	-	-
GoldFields	9.96e-05	-0.37	1.27	0.77	0.10	-0.34	-0.55
MurrayRob	1.54e-03	-0.09	0.76	0.00	-0.94	-	-
PicknPay	6.57e-04	-1.11	0.48	0.73	-	-	-

Table B.2: Shares - Estimates of the MA parameters

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
ABSA	-1.13	0.57	-0.72	0.62	0.12	-0.08
FirstRand	-0.29	-0.21	0.40	0.63	-0.68	-
Std Bank	-0.03	0.13	-0.07	-0.67	0.41	-
Nedbank	-1.75	1.71	-0.65	-	-	-
GoldFields	0.40	-1.31	-0.84	-0.05	0.38	0.52
MurrayRob	0.15	-0.76	-0.06	0.92	0.06	-
PicknPay	1.05	-0.61	-0.79	-	-	-

B.1.2 Probabilities Associated with the t-Statistics

Table B.3: Shares - Probabilities Associated with the constants and AR parameters

	c	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
ABSA	0.08	0.00	0.00	0.00	0.00	-	-
FirstRand	0.22	0.00	0.05	0.00	0.00	0.00	0.00
Std Bank	0.04	0.83	0.02	0.93	0.00	0.00	-
Nedbank	0.74	0.00	0.00	0.00	0.00	-	-
GoldFields	0.15	0.15	0.00	0.10	0.77	0.16	0.00
MurrayRob	0.02	0.01	0.00	0.91	0.00	-	-
PicknPay	0.05	0.00	0.00	0.00	-	-	-

Table B.4: Shares - Probabilities Associated with the MA parameters

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
ABSA	0.00	0.00	0.00	0.00	0.00	0.00
FirstRand	0.00	0.00	0.00	0.00	0.00	-
Std Bank	0.90	0.14	0.30	0.00	0.01	-
Nedbank	0.00	0.00	0.00	-	-	-
GoldFields	0.12	0.00	0.08	0.88	0.12	0.00
MurrayRob	0.00	0.00	0.12	0.00	0.00	-
PicknPay	0.00	0.00	0.00	-	-	-

B.2 Exchange Rates - ARMA(p,q)

B.2.1 Parameters

Table B.5: Exchange Rates - Estimates of the constants and AR parameters

	c	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
R/£	1.77e-04	-0.11	-0.57	-	-	-
R/€	2.23e-04	-	-	-	-	-
R/US$	5.62e-04	-1.94	-1.00	-	-	-
R/Aus$	2.02e-04	-0.09	-0.64	0.56	-	-
R/Yen	1.55e-04	0.27	0.41	-0.71	-0.41	0.57

Table B.6: Exchange Rates - Estimates of the MA parameters

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
R/£	0.08	0.55	-0.04	0.01	-0.05	0.07
R/€	-0.02	-0.02	-0.03	0.01	-0.03	0.07
R/US$	1.90	0.92	-0.06	-0.02	-	-
R/Aus$	0.05	0.62	-0.59	-	-	-
R/Yen	-0.31	-0.41	0.68	0.42	-0.61	-

B.2.2 Probabilities Associated with the t-Statistics

Table B.7: Exchange Rates - Probabilities Associated with the constants and AR parameters

	c	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
R/£	0.64	0.42	0.00	-	-	-
R/€	0.33	-	-	-	-	-
R/US$	0.51	0.00	0.00	-	-	-
R/Aus$	0.42	0.70	0.00	0.01	-	-
R/Yen	0.49	0.04	0.00	0.00	0.00	0.00

Table B.8: Exchange Rates - Probabilities Associated with the MA parameters

	1	2	3	4	5	6
R/£	0.58	0.00	0.02	0.51	0.00	0.00
R/€	0.14	0.16	0.06	0.36	0.10	0.00
R/US$	0.00	0.00	0.03	0.18	-	-
R/Aus$	0.82	0.00	0.01	-	-	-
R/Yen	0.01	0.00	0.00	0.00	0.00	-

B.3 Shares - VAR(6)

B.3.1 Parameters

$$\hat{\boldsymbol{C}} = \begin{pmatrix} 0.0009 \\ 0.0005 \\ 0.0010 \\ 0.0001 \\ 0.0008 \\ 0.0014 \\ 0.0009 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\mathbf{1}} = \begin{pmatrix} 0.0186 & 0.0406 & 0.0593 & 0.0444 & -0.0300 & -0.0279 & 0.0078 \\ 0.0687 & -0.0406 & 0.0499 & 0.0397 & -0.0105 & -0.0138 & 0.0002 \\ 0.0996 & 0.0402 & -0.0654 & 0.0027 & -0.0081 & -0.0041 & -0.0073 \\ 0.0541 & 0.0225 & 0.0475 & 0.0256 & -0.0268 & -0.0218 & 0.0075 \\ -0.0238 & 0.0616 & 0.0081 & -0.0452 & 0.0405 & 0.0241 & -0.0329 \\ 0.0584 & 0.0170 & -0.0211 & 0.0603 & 0.0164 & 0.0289 & 0.0242 \\ 0.0162 & 0.0424 & 0.0176 & 0.0120 & -0.0129 & 0.0019 & -0.0778 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\mathbf{2}} = \begin{pmatrix} -0.0720 & 0.0336 & -0.0176 & -0.0032 & 0.0225 & -0.0367 & -0.0009 \\ 0.0253 & -0.1052 & -0.0233 & 0.0214 & 0.0497 & 0.0018 & -0.0078 \\ 0.0131 & 0.0102 & -0.1071 & -0.0111 & 0.0397 & 0.0113 & -0.0230 \\ -0.0018 & 0.0184 & -0.0100 & -0.0711 & 0.0374 & -0.0046 & -0.0486 \\ -0.0518 & -0.0362 & 0.0293 & 0.1298 & -0.0443 & -0.0180 & 0.0274 \\ -0.0205 & 0.0283 & -0.0726 & 0.0105 & -0.0043 & -0.0083 & 0.0002 \\ 0.0106 & 0.0299 & -0.0323 & 0.0243 & 0.0278 & -0.0056 & -0.1004 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\mathbf{3}} = \begin{pmatrix} -0.0869 & -0.0155 & 0.0513 & 0.0073 & -0.0323 & 0.0332 & -0.0113 \\ 0.0295 & -0.0577 & -0.0375 & -0.0264 & -0.0179 & 0.0286 & 0.0282 \\ -0.0474 & 0.0319 & -0.0556 & -0.0188 & -0.0107 & 0.0323 & 0.0012 \\ -0.0365 & 0.0070 & 0.0510 & -0.0599 & -0.0228 & 0.0104 & -0.0088 \\ 0.0375 & -0.0266 & 0.0347 & -0.0559 & -0.0079 & 0.0471 & -0.0316 \\ 0.0325 & 0.0218 & -0.0232 & 0.0313 & 0.0046 & -0.0290 & -0.0374 \\ -0.0229 & 0.0039 & 0.0251 & 0.0175 & 0.0043 & -0.0245 & -0.0169 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\mathbf{4}} = \begin{pmatrix} -0.0384 & -0.0136 & 0.0213 & 0.0105 & -0.0001 & -0.0136 & 0.0124 \\ 0.0648 & -0.0875 & -0.0100 & 0.0017 & 0.0145 & -0.0299 & 0.0460 \\ -0.0182 & 0.0175 & -0.0689 & 0.0114 & -0.0028 & 0.0006 & 0.0089 \\ -0.0053 & 0.0331 & -0.0109 & -0.0594 & -0.0079 & -0.0106 & 0.0012 \\ 0.0392 & -0.0099 & -0.0263 & 0.0375 & 0.0238 & -0.0191 & 0.0394 \\ -0.0198 & -0.0783 & -0.0030 & 0.0240 & -0.0011 & -0.0090 & -0.0046 \\ -0.0037 & -0.0102 & 0.0021 & -0.0187 & -0.0017 & 0.0176 & -0.0427 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\mathbf{5}} = \begin{pmatrix} 0.0149 & 0.0234 & -0.0138 & -0.0220 & 0.0105 & 0.0087 & -0.0202 \\ 0.0201 & -0.0319 & -0.0226 & -0.0058 & 0.0072 & 0.0092 & -0.0096 \\ 0.0311 & 0.0213 & -0.0494 & -0.0179 & 0.0120 & -0.0043 & -0.0032 \\ 0.0007 & 0.0212 & 0.0011 & -0.0364 & 0.0101 & -0.0244 & 0.0260 \\ -0.0889 & -0.0160 & 0.0520 & 0.0912 & -0.0091 & -0.0308 & 0.0299 \\ 0.0121 & 0.0133 & 0.0043 & 0.0090 & -0.0345 & -0.0031 & -0.0165 \\ 0.0073 & -0.0017 & -0.0275 & 0.0124 & 0.0051 & 0.0153 & -0.0072 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{6}} = \begin{pmatrix} -0.0639 & 0.0851 & -0.0563 & -0.0269 & -0.0040 & -0.0175 & -0.0052 \\ 0.0242 & 0.0385 & -0.0535 & -0.0377 & 0.0049 & -0.0229 & 0.0015 \\ -0.0002 & 0.0440 & -0.0634 & -0.0660 & -0.0081 & -0.0099 & 0.0136 \\ -0.0133 & 0.0524 & -0.0121 & -0.0499 & 0.0014 & 0.0011 & 0.0063 \\ 0.0704 & 0.0074 & 0.0078 & 0.0008 & -0.0118 & -0.0442 & 0.0286 \\ 0.0316 & 0.0059 & 0.0168 & -0.0321 & -0.0017 & -0.0290 & -0.0265 \\ 0.0131 & 0.0385 & -0.0201 & -0.0573 & -0.0067 & -0.0086 & -0.0387 \end{pmatrix}$$

B.3.2 Probabilities Associated with the t-Statistics

$$\hat{\boldsymbol{C}}Prob = \begin{pmatrix} 0.05 \\ 0.23 \\ 0.02 \\ 0.79 \\ 0.23 \\ 0.01 \\ 0.02 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{1}}Prob = \begin{pmatrix} 0.49 & 0.16 & 0.04 & 0.09 & 0.02 & 0.13 & 0.74 \\ 0.01 & 0.17 & 0.08 & 0.14 & 0.42 & 0.46 & 0.99 \\ 0.00 & 0.17 & 0.02 & 0.92 & 0.54 & 0.82 & 0.76 \\ 0.04 & 0.44 & 0.10 & 0.34 & 0.04 & 0.24 & 0.75 \\ 0.57 & 0.17 & 0.85 & 0.27 & 0.04 & 0.40 & 0.36 \\ 0.06 & 0.61 & 0.52 & 0.05 & 0.27 & 0.17 & 0.37 \\ 0.51 & 0.11 & 0.50 & 0.62 & 0.28 & 0.91 & 0.00 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{2}}Prob = \begin{pmatrix} 0.01 & 0.25 & 0.54 & 0.90 & 0.08 & 0.05 & 0.97 \\ 0.35 & 0.00 & 0.42 & 0.42 & 0.00 & 0.92 & 0.74 \\ 0.63 & 0.73 & 0.00 & 0.68 & 0.00 & 0.54 & 0.33 \\ 0.95 & 0.53 & 0.73 & 0.01 & 0.00 & 0.80 & 0.04 \\ 0.21 & 0.42 & 0.51 & 0.00 & 0.03 & 0.53 & 0.45 \\ 0.51 & 0.40 & 0.03 & 0.73 & 0.77 & 0.70 & 0.99 \\ 0.67 & 0.26 & 0.22 & 0.32 & 0.02 & 0.74 & 0.00 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{3}}Prob = \begin{pmatrix} 0.00 & 0.60 & 0.08 & 0.78 & 0.01 & 0.07 & 0.63 \\ 0.28 & 0.05 & 0.20 & 0.32 & 0.17 & 0.12 & 0.23 \\ 0.08 & 0.28 & 0.06 & 0.48 & 0.42 & 0.08 & 0.96 \\ 0.18 & 0.81 & 0.08 & 0.02 & 0.08 & 0.57 & 0.71 \\ 0.37 & 0.55 & 0.44 & 0.17 & 0.70 & 0.10 & 0.39 \\ 0.29 & 0.51 & 0.48 & 0.30 & 0.76 & 0.17 & 0.17 \\ 0.36 & 0.88 & 0.34 & 0.47 & 0.72 & 0.15 & 0.43 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{4}} Prob = \begin{pmatrix} 0.16 & 0.64 & 0.46 & 0.69 & 0.99 & 0.46 & 0.60 \\ 0.02 & 0.00 & 0.73 & 0.95 & 0.27 & 0.11 & 0.05 \\ 0.50 & 0.55 & 0.02 & 0.67 & 0.83 & 0.98 & 0.71 \\ 0.85 & 0.26 & 0.71 & 0.03 & 0.55 & 0.57 & 0.96 \\ 0.35 & 0.83 & 0.56 & 0.36 & 0.24 & 0.50 & 0.28 \\ 0.52 & 0.02 & 0.93 & 0.43 & 0.94 & 0.67 & 0.87 \\ 0.88 & 0.70 & 0.94 & 0.44 & 0.89 & 0.30 & 0.05 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{5}} Prob = \begin{pmatrix} 0.58 & 0.42 & 0.63 & 0.41 & 0.42 & 0.64 & 0.39 \\ 0.46 & 0.27 & 0.44 & 0.83 & 0.59 & 0.62 & 0.68 \\ 0.25 & 0.47 & 0.09 & 0.50 & 0.36 & 0.82 & 0.89 \\ 0.98 & 0.46 & 0.97 & 0.17 & 0.44 & 0.19 & 0.27 \\ 0.03 & 0.72 & 0.24 & 0.03 & 0.65 & 0.28 & 0.41 \\ 0.69 & 0.69 & 0.90 & 0.77 & 0.02 & 0.88 & 0.54 \\ 0.77 & 0.95 & 0.30 & 0.61 & 0.67 & 0.37 & 0.74 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{6}} Prob = \begin{pmatrix} 0.02 & 0.00 & 0.05 & 0.31 & 0.76 & 0.34 & 0.82 \\ 0.37 & 0.19 & 0.06 & 0.16 & 0.71 & 0.22 & 0.95 \\ 0.99 & 0.13 & 0.03 & 0.01 & 0.54 & 0.59 & 0.57 \\ 0.62 & 0.07 & 0.67 & 0.06 & 0.91 & 0.95 & 0.79 \\ 0.09 & 0.87 & 0.86 & 0.98 & 0.56 & 0.12 & 0.43 \\ 0.30 & 0.86 & 0.61 & 0.29 & 0.91 & 0.17 & 0.33 \\ 0.59 & 0.15 & 0.45 & 0.02 & 0.58 & 0.61 & 0.07 \end{pmatrix}$$

B.4 Exchange Rates - VAR(6)

B.4.1 Parameters

$$\hat{\boldsymbol{C}} = \begin{pmatrix} 0.0001 \\ 0.0002 \\ 0.0001 \\ 0.0002 \\ 0.0002 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{1}} = \begin{pmatrix} 0.0163 & -0.0453 & -0.0734 & 0.0604 & 0.0106 \\ -0.0645 & 0.0228 & -0.0562 & 0.0505 & 0.0245 \\ -0.0373 & -0.0622 & -0.0916 & 0.0985 & 0.0662 \\ -0.0003 & -0.0334 & -0.0262 & 0.0059 & -0.0060 \\ -0.1421 & 0.0099 & -0.0794 & 0.1621 & 0.0287 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{2}} = \begin{pmatrix} 0.0723 & -0.0239 & -0.0684 & 0.0182 & -0.0205 \\ 0.1000 & -0.0788 & -0.0458 & 0.0272 & -0.0165 \\ 0.1348 & -0.0736 & -0.0794 & 0.0499 & -0.0158 \\ 0.0474 & 0.0071 & -0.0492 & 0.0236 & -0.0546 \\ 0.1428 & -0.1086 & -0.0723 & 0.0745 & -0.0209 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{3}} = \begin{pmatrix} 0.0073 & -0.0410 & 0.0131 & 0.0934 & -0.0717 \\ 0.0475 & -0.0452 & 0.0033 & 0.0615 & -0.0701 \\ 0.0260 & -0.0426 & -0.0070 & 0.0908 & -0.0578 \\ 0.0185 & -0.0292 & 0.0120 & 0.0666 & -0.0928 \\ 0.0863 & -0.1103 & -0.0016 & 0.0885 & -0.0677 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{4}} = \begin{pmatrix} 0.0724 & 0.0262 & -0.0813 & -0.0458 & 0.0368 \\ 0.0534 & 0.0152 & -0.0206 & -0.0214 & -0.0170 \\ 0.1241 & 0.0170 & -0.0533 & -0.0247 & -0.0171 \\ 0.0370 & 0.0727 & -0.0423 & -0.0443 & -0.0214 \\ 0.1696 & -0.0230 & -0.0623 & -0.0138 & -0.0374 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{5}} = \begin{pmatrix} -0.0980 & 0.0236 & 0.0840 & -0.0141 & -0.0207 \\ -0.0765 & 0.0323 & 0.0305 & 0.0084 & -0.0171 \\ -0.0870 & 0.0260 & 0.1044 & -0.0193 & -0.0695 \\ -0.0501 & 0.0227 & 0.0571 & -0.0043 & -0.0389 \\ -0.1142 & 0.0154 & 0.0650 & 0.0030 & -0.0274 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{6}} = \begin{pmatrix} 0.1335 & 0.0334 & -0.0142 & -0.0303 & -0.0652 \\ 0.0837 & 0.0551 & -0.0278 & 0.0071 & -0.0519 \\ 0.1358 & 0.0061 & -0.0097 & -0.0092 & -0.0606 \\ 0.0889 & 0.0124 & -0.0291 & -0.0010 & -0.0190 \\ 0.1475 & 0.0244 & -0.0142 & 0.0138 & -0.0836 \end{pmatrix}$$

B.4.2 Probabilities Associated with the t-Statistics

$$\hat{\boldsymbol{C}}Prob = \begin{pmatrix} 0.58 \\ 0.29 \\ 0.56 \\ 0.32 \\ 0.37 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{1}}Prob = \begin{pmatrix} 0.75 & 0.36 & 0.11 & 0.09 & 0.76 \\ 0.21 & 0.64 & 0.23 & 0.16 & 0.48 \\ 0.49 & 0.24 & 0.06 & 0.01 & 0.07 \\ 0.99 & 0.48 & 0.56 & 0.86 & 0.86 \\ 0.03 & 0.87 & 0.17 & 0.00 & 0.51 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\mathbf{2}} Prob = \begin{pmatrix} 0.16 & 0.63 & 0.14 & 0.61 & 0.56 \\ 0.05 & 0.11 & 0.32 & 0.45 & 0.64 \\ 0.01 & 0.16 & 0.11 & 0.19 & 0.67 \\ 0.33 & 0.88 & 0.27 & 0.49 & 0.10 \\ 0.03 & 0.08 & 0.21 & 0.10 & 0.63 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\mathbf{3}} Prob = \begin{pmatrix} 0.89 & 0.41 & 0.78 & 0.01 & 0.04 \\ 0.35 & 0.36 & 0.94 & 0.09 & 0.04 \\ 0.63 & 0.42 & 0.89 & 0.02 & 0.12 \\ 0.71 & 0.54 & 0.79 & 0.05 & 0.01 \\ 0.18 & 0.07 & 0.98 & 0.05 & 0.12 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\mathbf{4}} Prob = \begin{pmatrix} 0.16 & 0.60 & 0.08 & 0.20 & 0.29 \\ 0.30 & 0.76 & 0.66 & 0.55 & 0.63 \\ 0.02 & 0.75 & 0.28 & 0.52 & 0.65 \\ 0.45 & 0.13 & 0.34 & 0.20 & 0.52 \\ 0.01 & 0.71 & 0.28 & 0.76 & 0.39 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\mathbf{5}} Prob = \begin{pmatrix} 0.06 & 0.63 & 0.07 & 0.69 & 0.55 \\ 0.14 & 0.51 & 0.51 & 0.82 & 0.62 \\ 0.11 & 0.62 & 0.03 & 0.61 & 0.06 \\ 0.31 & 0.63 & 0.20 & 0.90 & 0.25 \\ 0.07 & 0.80 & 0.26 & 0.95 & 0.53 \end{pmatrix}$$

$$\hat{\boldsymbol{R}}_{\mathbf{6}} Prob = \begin{pmatrix} 0.01 & 0.50 & 0.76 & 0.40 & 0.06 \\ 0.10 & 0.27 & 0.55 & 0.84 & 0.14 \\ 0.01 & 0.91 & 0.84 & 0.81 & 0.10 \\ 0.07 & 0.79 & 0.51 & 0.98 & 0.57 \\ 0.02 & 0.69 & 0.81 & 0.76 & 0.06 \end{pmatrix}$$

Appendix C

Eigenvalues and Eigenvectors of $\hat{\boldsymbol{V}}$

C.1 Eigenvalues of $\hat{\boldsymbol{V}}$

The eigenvalues of $\hat{\boldsymbol{V}}$ are given below in Tables C.1 and C.2. The eigenvalues are the sample variances of the principal component scores. This was proved in section 3.1.

Table C.1: Shares - Eigenvalues of $\hat{\boldsymbol{V}}$

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	Number of the Principal Component						
		1	2	3	4	5	6	7
ARMA	$\boldsymbol{I}$	1.4122e-3	1.0850e-3	5.0916e-4	3.1557e-4	2.1861e-4	1.9855e-4	1.6623e-4
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	3.0436	1.0254	0.8583	0.8015	0.4786	0.4323	0.3602
VAR	$\boldsymbol{I}$	1.4008e-4	1.0724e-4	4.9896e-4	3.1304e-4	2.1327e-4	1.9120e-4	1.5897e-4
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	3.0651	1.0288	0.8546	0.8045	0.4736	0.4233	0.3501

Table C.2: Exchange Rates - Eigenvalues of $\hat{\boldsymbol{V}}$

$\hat{\boldsymbol{\mu}}_t$	$\boldsymbol{B}$	Number of the Principal Component				
		1	2	3	4	5
ARMA	$\boldsymbol{I}$	5.7228e-4	5.7579e-5	2.3564e-5	2.0101e-5	1.2223e-5
ARMA	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	4.1619	0.4139	0.1899	0.1400	0.0943
VAR	$\boldsymbol{I}$	5.6444e-4	5.6206e-5	2.2469e-5	1.9027e-5	1.1746e-5
VAR	$diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$	4.1774	0.4119	0.1835	0.1355	0.0918

C.2 $\boldsymbol{A}$ the Matrix of the Eigenvectors of $\hat{\boldsymbol{V}}$

Table C.3: Shares - $\boldsymbol{A}$ for ARMA(p,q) and $\boldsymbol{B} = \boldsymbol{I}$

$$\boldsymbol{A} = \begin{pmatrix} 0.4403 & -0.0708 & 0.1869 & -0.1497 & 0.6135 & -0.6006 & -0.0816 \\ 0.4749 & -0.0493 & 0.1410 & -0.0648 & 0.1140 & 0.4296 & 0.7419 \\ 0.4623 & -0.0664 & 0.1687 & -0.1437 & 0.0173 & 0.5427 & -0.6618 \\ 0.4385 & -0.0545 & 0.1462 & -0.1194 & -0.7808 & -0.3983 & 0.0280 \\ 0.1008 & 0.9923 & 0.0571 & -0.0414 & 0.0072 & -0.0062 & -0.0106 \\ 0.3195 & 0.0185 & -0.9443 & -0.0672 & 0.0213 & -0.0205 & -0.0210 \\ 0.2497 & 0.0130 & 0.0184 & 0.9655 & 0.0106 & -0.0345 & -0.0598 \end{pmatrix}$$

Table C.4: Shares - $\boldsymbol{A}$ for ARMA(p,q) and $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$

$$\boldsymbol{A} = \begin{pmatrix} 0.4493 & 0.1144 & -0.1581 & -0.1279 & 0.6003 & 0.6126 & -0.0901 \\ 0.4791 & 0.0432 & -0.1023 & -0.0866 & 0.1236 & -0.4146 & 0.7507 \\ 0.4707 & 0.0941 & -0.1330 & -0.1363 & 0.0339 & -0.5547 & -0.6515 \\ 0.4484 & 0.0677 & -0.1141 & -0.1237 & -0.7887 & 0.3789 & 0.0193 \\ 0.0219 & -0.9269 & -0.3081 & -0.2107 & 0.0167 & 0.0139 & -0.0221 \\ 0.2477 & -0.2540 & 0.9146 & -0.1871 & 0.0295 & 0.0296 & -0.0289 \\ 0.2902 & -0.2189 & 0.0483 & 0.9289 & 0.0039 & 0.0238 & -0.0463 \end{pmatrix}$$

Table C.5: Shares - $\boldsymbol{A}$ for VAR(p) and $\boldsymbol{B} = \boldsymbol{I}$

$$\boldsymbol{A} = \begin{pmatrix} 0.4403 & -0.0726 & 0.1824 & -0.1419 & 0.5682 & -0.6467 & -0.0797 \\ 0.4748 & -0.0492 & 0.1378 & -0.0698 & 0.1424 & 0.4163 & 0.7448 \\ 0.4633 & -0.0677 & 0.1737 & -0.1398 & 0.0642 & 0.5404 & -0.6594 \\ 0.4396 & -0.0582 & 0.1386 & -0.1288 & -0.8075 & -0.3391 & 0.0221 \\ 0.1022 & 0.9918 & 0.0595 & -0.0469 & 0.0052 & -0.0088 & -0.0111 \\ 0.3192 & 0.0209 & -0.9453 & -0.0497 & 0.0240 & -0.0163 & -0.0274 \\ 0.2459 & 0.0174 & 0.0346 & 0.9665 & -0.0031 & -0.0332 & -0.0524 \end{pmatrix}$$

Table C.6: Shares - $\boldsymbol{A}$ for VAR(p) and $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$

$$\boldsymbol{A} = \begin{pmatrix} 0.4509 & 0.1152 & 0.1360 & -0.1419 & 0.5722 & -0.6407 & -0.0832 \\ 0.4799 & 0.0419 & 0.0874 & -0.1023 & 0.1380 & 0.4124 & 0.7488 \\ 0.4716 & 0.0959 & 0.1215 & -0.1480 & 0.0610 & 0.5472 & -0.6547 \\ 0.4481 & 0.0729 & 0.0843 & -0.1395 & -0.8053 & -0.3443 & 0.0179 \\ 0.0224 & -0.9232 & 0.2692 & -0.2716 & 0.0119 & -0.0193 & -0.0228 \\ 0.2480 & -0.2505 & -0.9330 & -0.0468 & 0.0346 & -0.0226 & -0.0363 \\ 0.2850 & -0.2356 & 0.0947 & 0.9231 & -0.0034 & -0.0239 & -0.0406 \end{pmatrix}$$

Table C.7: Exchange Rates - $\boldsymbol{A}$ for ARMA(p,q) and $\boldsymbol{B} = \boldsymbol{I}$

$$\boldsymbol{A} = \begin{pmatrix} 0.4316 & -0.2296 & 0.5140 & 0.2007 & -0.6756 \\ 0.4311 & -0.2042 & 0.4988 & -0.3934 & 0.6074 \\ 0.4621 & 0.2056 & -0.1342 & 0.7754 & 0.3535 \\ 0.3640 & -0.6762 & -0.6265 & -0.1229 & -0.0508 \\ 0.5309 & 0.6372 & -0.2765 & -0.4344 & -0.2168 \end{pmatrix}$$

Table C.8: Exchange Rates - $\boldsymbol{A}$ for ARMA(p,q) and $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$

$$\boldsymbol{A} = \begin{pmatrix} 0.4625 & -0.1005 & 0.5012 & 0.4271 & -0.5851 \\ 0.4612 & -0.0919 & 0.5509 & -0.4376 & 0.5328 \\ 0.4562 & 0.3218 & -0.3120 & 0.6023 & 0.4778 \\ 0.4162 & -0.7418 & -0.5057 & -0.1262 & -0.0690 \\ 0.4381 & 0.5724 & -0.3037 & -0.4974 & -0.3752 \end{pmatrix}$$

Table C.9: Exchange Rates - $\boldsymbol{A}$ for VAR(p) and $\boldsymbol{B} = \boldsymbol{I}$

$$\boldsymbol{A} = \begin{pmatrix} 0.4325 & -0.2230 & 0.5057 & 0.1994 & -0.6839 \\ 0.4315 & -0.2114 & 0.4958 & -0.4270 & 0.5840 \\ 0.4615 & 0.2151 & -0.0927 & 0.7681 & 0.3771 \\ 0.3647 & -0.6773 & -0.6319 & -0.0857 & -0.0408 \\ 0.5298 & 0.6330 & -0.3009 & -0.4250 & -0.2177 \end{pmatrix}$$

Table C.10: Exchange Rates - $\boldsymbol{A}$ for VAR(p) and $\boldsymbol{B} = diag\{\hat{\phi}_{ii}^{-\frac{1}{2}}\}$

$$\boldsymbol{A} = \begin{pmatrix} 0.4628 & -0.0972 & 0.4885 & 0.4376 & -0.5885 \\ 0.4607 & -0.1014 & 0.5608 & -0.4533 & 0.5074 \\ 0.4563 & 0.3290 & -0.2917 & 0.5897 & 0.5007 \\ 0.4164 & -0.7399 & -0.5128 & -0.1124 & -0.0596 \\ 0.4382 & 0.5697 & -0.3144 & -0.4926 & -0.3767 \end{pmatrix}$$

Bibliography

C. O. Alexander. A primer on the orthogonal garch model. Technical report, University of Reading, 2000.

C. O. Alexander. Principal component models for generating large garch covariance matrices. *Review of Banking, Finance and Monetary Economics*, 31:337–359, 2003.

L. Bauwens, S. Laurent, and J. V. Rombouts. Multivariate garch models: A survey. *Journal of Applied Econometrics*, 21:79–109, 2006.

T. Bollerslev. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *Review of Economics and Statistics*, 72:498–505, 1988.

T. Bollerslev, R. Engle, and J. Wooldridge. A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96:116–131, 1988.

M. B. Bongers. Multivariate volatility modelling in modern finance. Master's thesis, University of Cape Town, 2008.

H. P. Boswijk and R. van der Weide. Wake me up before you go-garch. Technical Report TI 2006-079/4, Tinbergen Institute, 2006.

A. Chibumba and C. O. Alexander. Multivariate orthogonal factor garch. Technical report, University of Sussex, 1996.

Z. Ding. *Time Series Analysis of Speculative Returns*. PhD thesis, University of California, University of California, San Diego, 1994.

R. Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, 20:339–350, 2002.

R. Engle and K. Kroner. Multivariate simultaneous generalized arch. *Econometric Theory*, 11:122–150, 1995.

J. Fan, M. Wang, and Q. Yao. Modelling multivariate volatilities via conditionally uncorrelated components. *Journal of the Royal Statistical Society*, 70:679–702, 2008.

A. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, 1989.

J.P. Morgan and Reuters. Riskmetrics - technical document. Technical report, J.P. Morgan Bank, New York, 1996.

B. Manly. *Multivariate Statistical Methods: A Primer*. Chapman and Hall, third edition, 2005.

G. Reinsel. *Elements of Multivariate Time Series Analysis*. New York: Springer, second edition, 1997.

N. Shephard. Local scale models. *Journal of Econometrics*, 60:181–202, 1994.

R. Tsay. *Analysis of Financial Time Series*. Wiley, NY, second edition, 2005.

Y. K. Tse and A. K. C. Tsui. A note on diagnosing multivariate conditional heteroscedasticity models. *Journal of Time Series Analysis*, 20:351–362, 1999.

Y. K. Tse and A. K. C. Tsui. A multivariate garch model with time-varying correlations. *Journal of Business and Economic Statistics*, 20:351–362, 2002.

R. van der Weide. Go-garch: A multivariate generalized orthogonal garch model. *Journal of Applied Econometrics*, 17:549–564, 2002.

I. Vrontos, P. Dellaportas, and D. Politis. A full-factor multivariate garch model. *Journal of Applied Econometrics*, 6:312–334, 2002.